

# **Aalto-yliopiston teknillinen korkeakoulu**

Sähkötekniikan laitos

Elektroniikan, tietoliikenteen ja automaation tiedekunta

Matti Koivisto

## **Tuntimittausdatan käyttö sähkökuorman ennustamisessa**

Diplomityö, joka on jätetty opinnäytteenä tarkastettavaksi diplomi-insinöörin tutkintoa varten Espoossa 3.5.2010

Työn valvoja ja ohjaaja

Professori     Matti Lehtonen

AALTO-YLIOPISTON  
TEKNILLINEN KORKEAKOULUDIPLOMITYÖN  
TIIVISTELMÄ

Tekijä: Matti Koivisto	
Työn nimi: Tuntimittausdatan käyttö sähkökuorman ennustamisessa	Sivumäärä:
Päivämäärä: 3.5.2010	87
Elektroniikan, tietoliikenteen ja automaation tiedekunta Sähkötekniikan laitos Professori: S-18 Sähköverkot ja suurjännetekniikka	
Työn valvoja ja ohjaaja: Professori Matti Lehtonen	
<p>Tässä työssä tutkitaan sähkökuorman ennustamista regressionanalyysin avulla. Selittävinä tekijöinä käytetään lämpötilaa ja päivänpituutta. Vuoden mittainen tuntimitattua dataa sisältävä aineisto (1.7.2008–30.6.2009) saatiin Kainuun Energialta n. 1600 asiakkaalta ja se sisältää lähinnä kotitalouksien sähkönkulutusdataa; tehty ohjelma onkin pääasiassa suunniteltu kotitalouksien sähkönkulutuksen ennustamista varten.</p> <p>Ohjelmalla voi analysoida pienten asiakasryhmien summakulutusta, jolloin voidaan ennustaa esimerkiksi yhden jakelumuuntajan kulutus. Tämän lisäksi voidaan käsitellä suuremman asiakasjoukon keskiarvoa; tätä käytetään lähinnä ohjelman testaamiseen. Koska lämpötilan vaikutus sähkönkulutukseen on epälineaarinen, vuosi jaetaan päiväryhmiin jotka käsitellään erikseen. Ryhmän sisällä oletetaan lineaarinen riippuvuus.</p> <p>Regressioanalyysin antamat selittävien muuttujien kertoimet tarkistetaan automaattisesti ohjelmassa määriteltujen periaatteiden mukaisesti, jotta ennustemalli olisi järkevä. Tämän lisäksi lasketaan residuaalien hajonta, jolloin voidaan antaa ennusteen lisäksi haluttu luottamustaso. Myös luottamustason käyttöä varten tarvittavaa normaalijakaumaoletusta sekä oletettua residuaalien riippumaattomuutta selittävistä tekijöistä tutkitaan.</p> <p>Lopuksi verrataan ennustemallin antamia arvoja toteutuneeseen vuoden 2010 tammikuun kulutukseen. Tehtyjen testien perusteella ennuste toimii melko hyvin, mutta joitain kysymyksiä jäi vielä auki. Nämä liittyvät lähinnä tapaan käsitellä epälineaarisuus, sekä arvioon analysoitavan ryhmän minimiasiakasmäärästä.</p>	
Avainsanat: sähkökuorma, regressioanalyysi, ennustaminen, jakelumuuntajan kuormitus	

AALTO UNIVERSITY  
SCHOOL OF SCIENCE AND TECHNOLOGY

ABSTRACT OF THE  
MASTER'S THESIS

Author: Matti Koivisto	
Title: Hourly measured data in electrical load forecasting	Number of pages:
Date: 3.5.2010	87
Faculty of Electronics, Communications and Automation Department of Electrical Engineering Professorship: S-18 Power Systems and High Voltage Engineering	
Supervisor and instructor: Professor Matti Lehtonen	
<p>The use of regression analysis for electrical load forecasting is investigated in this thesis. The explanatory variables are temperature and day length. The year long sample data containing the hourly measured consumption (1.7.2008-30.6.2009) of about 1,600 customers consists mainly of household electricity consumption data, and the created program is designed mainly for forecasting household electricity consumption.</p> <p>The program can analyze the consumption of small groups, so we can predict, for example, the consumption of a distribution transformer. In addition, we can analyze the mean of a larger number of customers; this is used mainly for testing the program. Since the effect of temperature on electricity consumption is non-linear, the year has been divided into day groups, which are treated separately. Linear dependence is assumed within a group.</p> <p>Regression coefficients for the explanatory variables are examined automatically in accordance with the defined principles, in order to get a reasonable forecast model. In addition, the standard deviation of the calculated residuals is calculated, so that we can present the desired confidence level. The required normal distribution assumption and the expected independence of the residuals from the explanatory variables have also been studied.</p> <p>Finally, the forecast model is tested against the measured values from January 2010. The forecast model is fairly successful according to these tests, but some questions remain open. These relate mainly to the way to address non-linearity, as well as to a suitable estimate for the minimum number of customers in a group to be analyzed.</p>	
Keywords: electrical load, regression analysis, prediction, load of distribution transformer	

## Alkulause

Diplomityö on tehty Aalto-yliopiston teknillisen korkeakoulun elektroniikan, tietoliikenteen ja automaation tiedekunnassa opinnäytteeksi sähkötekniikan laitokselle. Haluan kiittää työn valvojaa ja ohjaajaa professori Matti Lehosta mielenkiintoisen ja ajankohtaisen diplomityöaiheen järjestämisestä, opastuksesta sekä hyvän yleisnäkemyksen tarjoamisesta nykyisin käytössä olevista ennustemenetelmistä.

Kiitos E.On:lle (entinen Kainuun Energia) työn lähtökohtana olleen sähkönmittausdatan antamisesta, sekä kaikille niille ihmisille jotka allekirjoituksellaan hyväksyivät sähkönmittausdatansa käytön tämän tutkimustyön lähtödatana. Kiitos myös Helsingin Energialle, Vantaan Energialle ja Teklalle näkemyksistä niistä tarpeista ja rajoituksista joita verkostolaskennassa sekä kulutuksen ennustamisessa on tuntimitatun datan käyttöä koskien. Erityiskiitos Vantaan Energialle diplomityötä tekeväälle tekniikan ylioppilas Timo Kaartiolle näkemyksistä pitkän aikavälin ennustamisen tarpeista koskien tuntimitatun datan analysointia.

Haluan kiittää avusta myös koko laboratoriomme henkilöstöä; erityisesti DI Jarno Lamposta, DI Anssi Aholaa, DI Paavo Tammea ja tekniikan tohtori John Millaria. Erityiskiitos tekniikan ylioppilas Merkebu Degefalle mittausdatan käsittelystä sekä avusta datan analysoimisessa.

Kiitos myös vanhemmilleni opintojeni tukemisesta.

Espoossa 3. toukokuuta 2010

Matti Koivisto

# Sisällysluettelo

1	Työn lähtökohdat.....	1
1.1	Kotitalousasiakkaiden kulutuksen ennustaminen.....	2
1.2	Työssä käytetty data.....	3
1.3	Työn tavoite.....	4
2	Analysoitavat aikasarjat .....	6
2.1	Tutkittavat asiakastyypit .....	6
2.2	Asiakastyypien keskiarvojen analysointi .....	9
2.3	Pienen asiakasryhmän summan analysointi .....	9
3	Vuoden päivien jakaminen ryhmiin .....	12
3.1	Jako arki-, aatto- ja pyhäpäiviin .....	12
3.2	Vuoden päivien jakaminen lämpötila-alueisiin.....	14
3.3	Kahden lämpötilamuuttujan käyttö.....	16
3.4	Lämpötilan vaikutuksen hitaus.....	17
4	Hajonnan tutkiminen .....	19
4.1	Normaalijakauman oletaminen.....	20
4.2	Päiväenergioiden hajonta .....	20
4.3	Tunneittainen hajonta .....	22
4.4	Usean asiakkaan summan varianssi.....	25
5	Ennusteen muodostaminen .....	29
5.1	Päivänpituuden laskeminen.....	32
5.2	Kahden muuttujan regressioanalyysi .....	34
5.3	Ohjelman toiminnasta .....	37
5.4	Regressiotulosten automaattinen tarkistus .....	41
6	Ennusteen analysointi.....	43
6.1	Eri lämmitystyyppien päivärakenteet .....	43
6.2	Eri lämmitystyyppien huippukulutuksen ennustaminen .....	48
6.3	Muuntajan summan ennustaminen .....	51

7	Ennuste vuoden 2010 tammikuulle .....	56
7.1	Eri lämmitystyypit tammikuussa 2010 .....	56
7.2	Muuntajan summakuormitus .....	61
7.3	Varaavan sähkölämmittäjän kulutusennuste .....	63
7.4	Ilmalämpöpumppua käyttävä suora sähkölämmittäjä .....	66
8	Yhteenveto ja jatkotutkimusehdotukset .....	69
	Lähdeluettelo .....	71
	Liite 1: Varaavan sähkölämmittäjän jako kolmeen lämpötila-alueeseen .....	73
	Liite 2: Päivänpituuden vaikutus öljylämmittäjän kulutukseen .....	74
	Liite 3: Suoran sähkölämmittäjän mitattu ja ennustettu kulutus .....	75
	Liite 4: Öljylämmittäjän mitattu ja ennustettu kulutus .....	76
	Liite 5: Varaavan sähkölämmittäjän mitattu ja ennustettu kulutus kun jako kolmeen lämpötila-alueeseen .....	77
	Liite 6: Suoran sähkölämmittäjän kylmien arkipäivien rakenne .....	78
	Liite 7: Kaukolämpöasiakkaan kylmien arkipäivien rakenne .....	79
	Liite 8: Varaavan sähkölämmittäjän kylmien pyhäpäivien rakenne .....	80
	Liite 9: Öljylämmittäjän kylmien aattopäivien rakenne .....	81
	Liite 10: Suoran sähkölämmittäjän kulutushuiput 2008-2009 .....	82
	Liite 11: Kaukolämpöasiakkaan kulutushuiput 2008-2009 .....	83
	Liite 12: Kaukolämpöasiakas 21-31.1.2010 kun jako kahteen lämpötila-alueeseen .....	84
	Liite 13: Öljylämmittäjä 11-20.1.2010 kun jako kahteen lämpötila-alueeseen .....	85
	Liite 14: Varaava sähkölämmittäjä 11-31.1.2010 kun jako kolmeen lämpötila-alueeseen .....	86
	Liite 15: Ilmalämpöpumppua käyttävä suora sähkölämmittäjä 1-20.1.2010 .....	87

# 1 Työn lähtökohdat

Sähköverkon täytyy olla tasapainossa tuotannon ja kulutuksen suhteen. Siksi on tärkeää pystyä ennustamaan tulevaisuuden sähkönkulutus. Lyhyen aikavälin ennuste tarvitaan verkon käyttöä, keskipitkän aikavälin ennuste vuoden kulutushuippujen ennustamista ja pitkän aikavälin ennuste tulevaisuuden verkon ja sähköntuotannon suunnittelemista varten. Sähkönkulutusarvio on perusta sähköverkon kehittämiselle ja sähkön tuotannon suunnittelemiselle.

Tässä työssä tutkitaan mahdollisuutta käyttää tuntimitattua dataa erityyppisten ja erikokoisten asiakasryhmien kulutuksen ennustamiseksi. Suomessa asennetaan koko ajan enemmän tunneittaisen sähkönkulutuksen mittaavia mittareita kotitalouksiin ja 2013 suurin osa Suomen sähkönkulutuksesta pitäisi olla tuntimitattua [12]. Tällä hetkellä kotitalouksien kulutus on arvioitu erityyppisten asiakkaiden keskiarvojen perusteella, mutta tulevaisuudessa on mahdollisuus paljon tarkempaan sähkönkulutuksen ennustamiseen.

Työssä keskitytään jakelumuuntajatasoon, ja lähinnä kotitalouksiin koska niistä saatiin kattava määrä dataa. Yksittäisen asiakkaan sähkönkulutuksen ennustaminen on periaatteessa mahdollista, mutta hyvin hankalaa suuren hajonnan takia. Tavoite on pystyä ennustamaan noin kymmenen tai parinkymmenen asiakkaan ryhmiä, mutta tarvittaessa voidaan ennustaa myös isompia ryhmiä. Näin voidaan esimerkiksi laskea sähkönkulutusennuste tietyn jakelumuuntajan syöttämälle asiakasryhmälle, jolloin voidaan arvioida muuntajan kuormitus eri tilanteissa. Haja-asutusalueella yksi jakelumuuntaja voi kuitenkin syöttää vain yhtä asiakasta joten ei voida luvata että aina pystytään arvioimaan kunkin jakelumuuntajan kulutus. Tässä työssä on tavoite myös antaa arvio hyvän ennusteen tarvitsemasta minimiasiakasmäärästä.

Työssä on otettu lähtökohdaksi yhden vuoden tuntimitattu data, koska yli vuoden mittaista jaksoa tuntimitattua dataa ei ole vielä saatavilla suurelta joukolta asiakkaita. Laskutusta varten ainakin yhden vuoden data tullaan säilyttämään jokaiselta asiakkaalta, mutta ilmeisesti tuntimitattua sähkönkulutusdataa koskevan asetuksen mukaisesti sähköyhtiöiden tulee säilyttää dataa kuusi vuotta. Kappaleessa 8 arvioidaan mahdollisuuksia käyttää dataa pidemmältä ajalta.

Työssä tutkitaan tärkeimpien sähkönkulutukseen vaikuttavien tekijöiden huomioon ottamista mahdollisimman automaattisesti. Suomessa tärkein vaikuttava tekijä on lämpötila, mutta myös päivänpituus voi joillain asiakkailla olla tärkeä sähkönkulutuksen vuosivaihtelua selittävä

ulkoinen tekijä [5]. Näiden ulkoisten tekijöiden lisäksi tärkeitä huomioitavia asioita ovat eri päivätyyppien, eli arki-, aatto- ja pyhäpäivien erot sekä päivän sisällä eri tuntien erot.

Eri lämmitystyyppien erot ovat suuria. Sähkön käyttäminen lämmitykseen on suuri sähkön käyttöä lisäävä tekijä. Jotkut asiakkaat voivat käyttää sähköä selvästi enemmän yöllä, jos asiakas on solminut yö sähkönsopimuksen ja näin saa yöllä sähkön halvemmalla. Tässä työssä keskitytään pelkästään kotitalouksien sähkönkulutuksen ennustamiseen koska tämä on tärkein uusi tuntimittauksen piiriin tuleva kuluttajaryhmä, eikä toimisto- ja pienyrityksistä ollut tarpeeksi dataa saatavilla. Arkipäivät kuitenkin erotetaan lomapäivistä, joten esitetty automaattinen malli voisi sopia myös esimerkiksi pienten yritysten sähkönkulutuksen ennustamiseen. Suurten sähkönkäyttäjien kulutus on ollut jo pitkään saatavilla tuntimitattuna: niitä ei käsitellä tässä työssä.

## 1.1 Kotitalousasiakkaiden kulutuksen ennustaminen

Nykyään asiakkaiden kotitalouksien sähkönkulutuksesta tiedetään yleensä vain vuosienergia. Tämän avulla voidaan ennustaa myös tietyn hetken teho tai huipputeho esimerkiksi viitteessä [10] kappaleessa 3.9 esitetyillä tavoilla. Tällöin on kuitenkin tärkeää tietää asiakkaan tyyppi. Tässä työssä esitetty mallinnus on periaatteeltaan samantyyppinen kuin viitteen [10] kappaleessa 3.9.2 esitetty kuormitusmalleihin perustuva menetelmä, mutta tässä kuormitusmalli muodostetaan analysoimalla vuoden mittainen tuntimitattu data ilman mitään tietoa asiakkaiden tyypistä.

Sähkönkulutuksen mallintamisesta on tehty paljon tutkimusta. Näistä useat ovat keskittyneet sähkökauppaan liittyvään kulutuksen mallintamiseen, kuten esimerkiksi viitteessä [11] esitetty SARIMAX-malli. Sähkökauppaa varten tarvitaan erittäin tarkka malli joka ennustaa usein vain lyhyeksi aikaa eteenpäin. Tällainen malli koskee usein suurta massaa, esimerkiksi koko Suomen kulutusta. Tässä työssä lähestytään sähkönkulutuksen ennustamista hieman eri näkökulmasta. Mallin ei tarvitse olla yhtä tarkka mutta ennustettaessa halutaan saada malli koko vuoden kulutukselle ja varsinkin huippukulutuksen ennustaminen on tärkeää. Tässä halutaan myös käsitellä melko pienien asiakasryhmien kulutusta, joten satunnaisvaihtelun määrä tulee olemaan suuri.

Toinen määräävä tekijä on tavoite pitää malli yksinkertaisena ja helposti ymmärrettävänä, joten halutaan välttää ns. ”black box”- tyyppistä mallia jossa syötetyn datan ja ulostulevan ennusteen



välissä tapahtuva analyysi ei välttämättä ole kovin selkeästi ja suoraviivaisesti määritelty. Tämän lähestymistavan toivotaan johtavan robustimpaan ja luotettavampaan, tosin esimerkiksi neuroverkkomalleja epätarkempaan, malliin.

## 1.2 Työssä käytetty data

Työssä käytetty mittausdata saatiin Kainuun alueelta Kainuun Energialta. Noin 1600 asiakasta vastasi lähetettyyn kyselyyn ja antoi luvan mittausdatansa käyttöön tutkimuksessa. Asiakkaat myös vastasivat lähetettyyn kyselyyn jossa kysyttiin muun muassa lämmitystyyppi. Näin oli mahdollista jakaa asiakkaat lämmitystyyppien mukaisesti ryhmiin. Mittausdataa saatiin vuoden mittainen jakso ajalta 1.7.2008–30.6.2009. Sekä tunneittainen mittausdata että kyselyjen tiedot tallennettiin mySql-tietokantaan. Näin voitiin hakea tietyn tyyppisiä asiakasryhmiä analyysiä varten. Analyysiä varten tehtiin C++ ohjelma. Sähkömittausdatan lisäksi ilmatieteen laitokselta saatiin samalta ajalta tunneittaista lämpötiladataa Kainuun lentokentältä. Samaa lämpötiladataa käytettiin kaikkien asiakkaiden ennusteen tekemisessä.

Työn tekemisen loppupuolella saatiin kulutus- ja lämpötiladataa myös tammikuulta 2010. Tätä dataa ehdittiin testata jonkin verran ja saadut tulokset on esitelty kappaleessa 7. Data saatiin noin 80 % niistä asiakkaista joilta oli dataa ajalta 1.7.2008–30.6.2009. Tammikuu 2010 yli hyvin kylmä verrattuna vuosien 2008–2009 talveen joten se tarjosi hyvän mahdollisuuden testata ennusteen toimivuutta talvella.

Tavoite on rakentaa mahdollisimman automaattinen järjestelmä jossa asiakkaista ei tarvitse etukäteen tietää mitään, mutta järjestelmän testaamista varten haluttiin selkeästi määritellyt asiakasryhmät jotta tuloksia voidaan verrata niihin oletuksiin joita esimerkiksi suoraa sähkölämmitystä käyttävistä asiakkaista on. Tämän takia asiakkaat jaettiin kyselyjen vastausten perusteella ryhmiin, joista kerrotaan enemmän kappaleessa 2.1. Tiettyjen asiakastyypin käyttäytymisen lisäksi haluttiin testata myös pienen satunnaisesti valitun asiakasryhmän analyysin toimivuutta, joten tietokannasta valittiin satunnaisesti 5, 10, 30 ja 50 asiakkaan ryhmä. Nämä kuvaavat testissä jakelumuuntajan kuormaa ja ne käsitellään summana joten lopputulos on koko muuntajan kuorman käyttäytyminen vuoden aikana.

## 1.3 Työn tavoite

Tämän työn tavoite on tehdä ennuste pienen asiakasryhmän summateholle. Tämä ryhmä voi olla esimerkiksi muuntajan takana olevat asiakkaat jotka voivat olla hyvin samanlaisia tai täysin erilaisia. Tarkoitus on kokeilla minkä kokoisella ryhmällä analyysi voidaan tehdä. Käytettävissä on yhden vuoden tuntimitattu sähkönkulutusdata sekä samalta ajalta joko tunneittainen tai päiväkeskiarvoista koostuva lämpötiladata.

Analyysi tehdään C++ ohjelmalla ja analyysin tulee tapahtua nopeasti ja luotettavasti, jotta jakeluverkossa voidaan laskea suurten muuntajamäärien ennusteet automaattisesti. Näin saadaan tehonjakoa varten tarvittavat kulutusennusteet halutulla hetkellä. Mahdollinen käytännön ennustamistilanne voisi olla esimerkiksi seuraavanlainen: ohjelman käyttäjä haluaa tammikuun arkipäivän kello kahdentoista ennusteen jos ulkona on 25 astetta pakkasta. Tämän jälkeen ohjelma käy läpi kaikki määritellyt ryhmät, esimerkiksi jakelumuuntajat, ja asettaa niihin analyysin antaman kulutusarvon valitulla hetkellä. Ohjelman voi myös määritellä esimerkiksi etsimään tietyn alueen summakulutuksen huipun ja sen jälkeen asettamaan yksittäisille jakelumuuntajille kyseisen tunnin kulutuksen.

Pienien ryhmien summan lisäksi ohjelmassa on mahdollisuus tutkia suurten asiakasryhmien keskiarvoja. Tämä on lähinnä ohjelman testausta varten koska esimerkiksi sadan öljylämmittäjän ryhmän keskiarvon tulee noudattaa tiettyä loogista riippuvuutta esimerkiksi lämpötilasta. Keskiarvolle tehtyä ennustemallia tarkastelemalla voidaan arvioida pystyykö ohjelma käsittelemään esimerkiksi varaavan sähkölämmityksen päivärakennetta ja onko kaukolämmittäjän ennusteen vuosirakenne uskottava.

Lähtökohta on käyttää mahdollisimman vähän selittäviä tekijöitä ja olettaa aina lineaarinen yhteys selittävän ja selitettävän välillä. Tämä ei anna tarkinta mahdollista tulosta ja saattaa jossain tilanteessa antaa melko pahastikin vääriä tuloksia. Lineaarinen selittäminen on kuitenkin hyvin suoraviivaista ja helppoa, eikä se yleensä anna täysin epäloogisia tuloksia esimerkiksi laskennallisen ongelman takia; esimerkiksi eksponentiaaliseen yhtälöön perustuvassa ennusteessa riski karkeasti väärin tuloksiin on suurempi. Lineaarisen yhteyden olettamisen aiheuttamat ongelmat on yritetty ratkaista kun ne ovat tulleet eteen ja ne on esitetty tässä työssä.

Luotettavuutta ja selkeyttä painotetaan, jopa tarkkuuden kustannuksella jos nämä ovat ristiriidassa. Koska ohjelman on tarkoitus olla mahdollisimman automaattinen, ei mahdollisuutta

---

karkeisiin virheisiin haluta turhaan kasvattaa. Toisaalta pieni epätarkkuus on hyväksyttävää, koska pienen asiakasryhmän ennuste ei voi kuitenkaan olla äärimmäisen tarkka suuresta satunnaisvaihtelusta johtuen.

## 2 Analysoitavat aikasarjat

Tässä työssä analysoidaan aina vain yksi aikasarja, joka on vuoden mittainen ja sisältää jokaiselle tunnille kulutusarvon. Sarja sisältää siis 8760 tai 8784 mittausta. Aikasarja voi periaatteessa olla minkälainen tahansa eli esimerkiksi yksittäisen kesämökin kulutus voidaan analysoida. Ohjelmaa on kuitenkin testattu vain tietyillä sarjoilla jotka koostuvat yleensä useista asiakkaista jotka käyttävät sähköä koko vuoden ajan. Nämä sarjat voivat olla joko ryhmän asiakkaiden mittaussarjoista koostettuja keskiarvo- tai summasarjoja. Yksittäisistä asiakkaista on poistettu ne, joiden mittaussarja on selkeästi puutteellinen. Jos asiakkaan mittausdatassa on esimerkiksi yli kuukauden jakso ilman mitään sähkönkulutusta, on kyseinen asiakas poistettu ryhmästä.

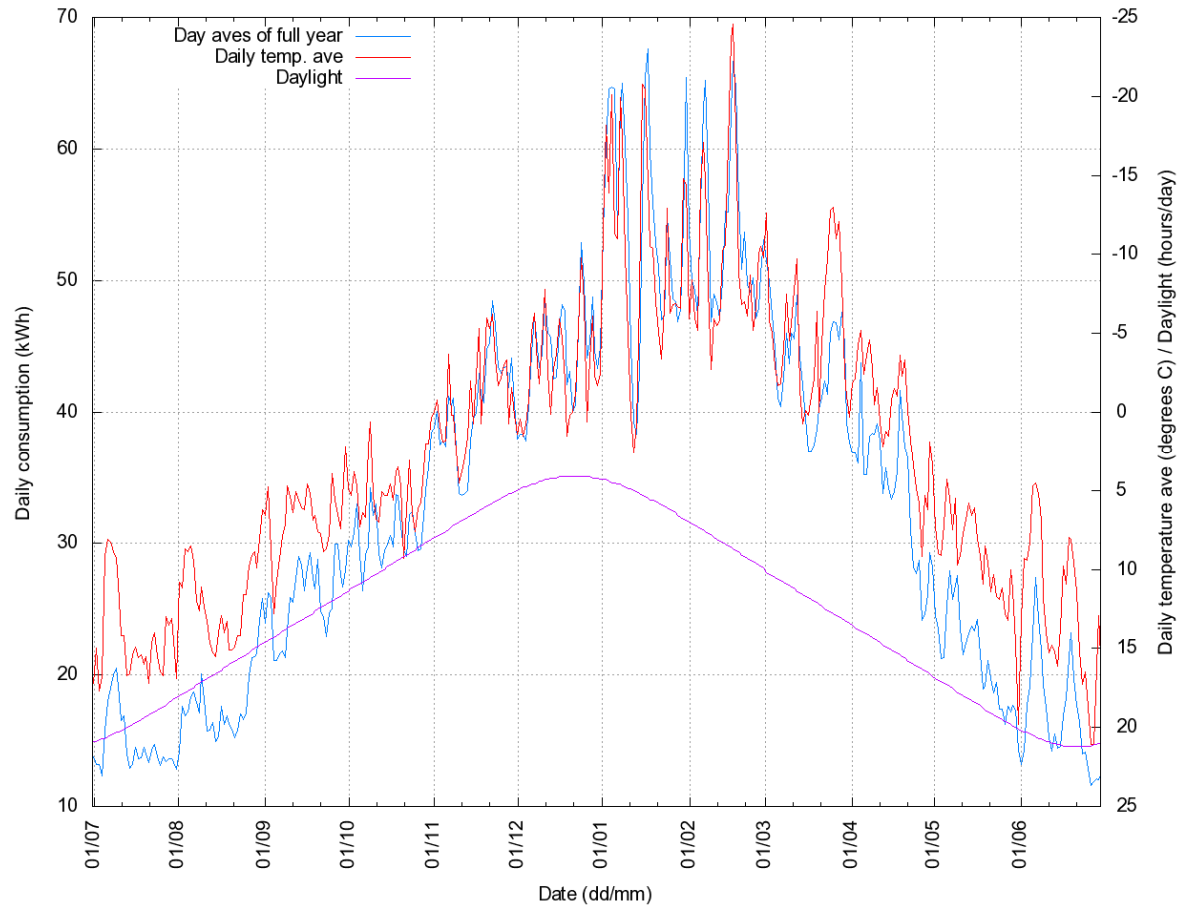
### 2.1 Tutkittavat asiakastyypit

Tässä työssä tutkitaan neljän eri lämmitystapaa käyttävän asiakastyypin sähkönkulutusta. Asiakasryhmien yksittäisten asiakkaiden samankaltaisuutta ei ole testattu, vaan on tyydytty siihen että ryhmän keskiarvo vastaa yleisellä tasolla asiakasryhmän oletettua käyttäytymistä. Tämä on todettu vain tarkastelemalla silmämääräisesti eri asiakastyypin rakennetta ja vertaamalla niitä toisiinsa.

Lämmitystyypeistä on tarkasteluun valittu neljä erilaista. Pelkkää suoraa sähkölämmitystä käyttäviä asiakkaita (direct electrical heating) on 159 kappaletta ja pelkkää varaavaa sähkölämmitystä (electrical storage heating) 23. Öljylämmitystä (oil heating) käyttäviä asiakkaista mukana on 77 ja kaukolämpöasiakkaista (district heating) 281. Tehdyn ohjelman tulostamat kuvat ovat englanniksi joten niissä esiintyy edellä suluissa näkyviä termejä käyrien selitysteksteissä. Asiakasmääriltään muut lämmitystyyppit sisältävät tarpeeksi asiakkaita, mutta varaavaa sähkölämmitystä käyttäviä asiakkaita on melko vähän. Tämä on valitettavaa koska asiakastyypinä se on hankala mutta tärkeä ennustaa. Näinkin pienen asiakasryhmän keskiarvo tuntui kuitenkin noudattavan loogista päivärytmiä ja ennusteen muodostaminen onnistui.

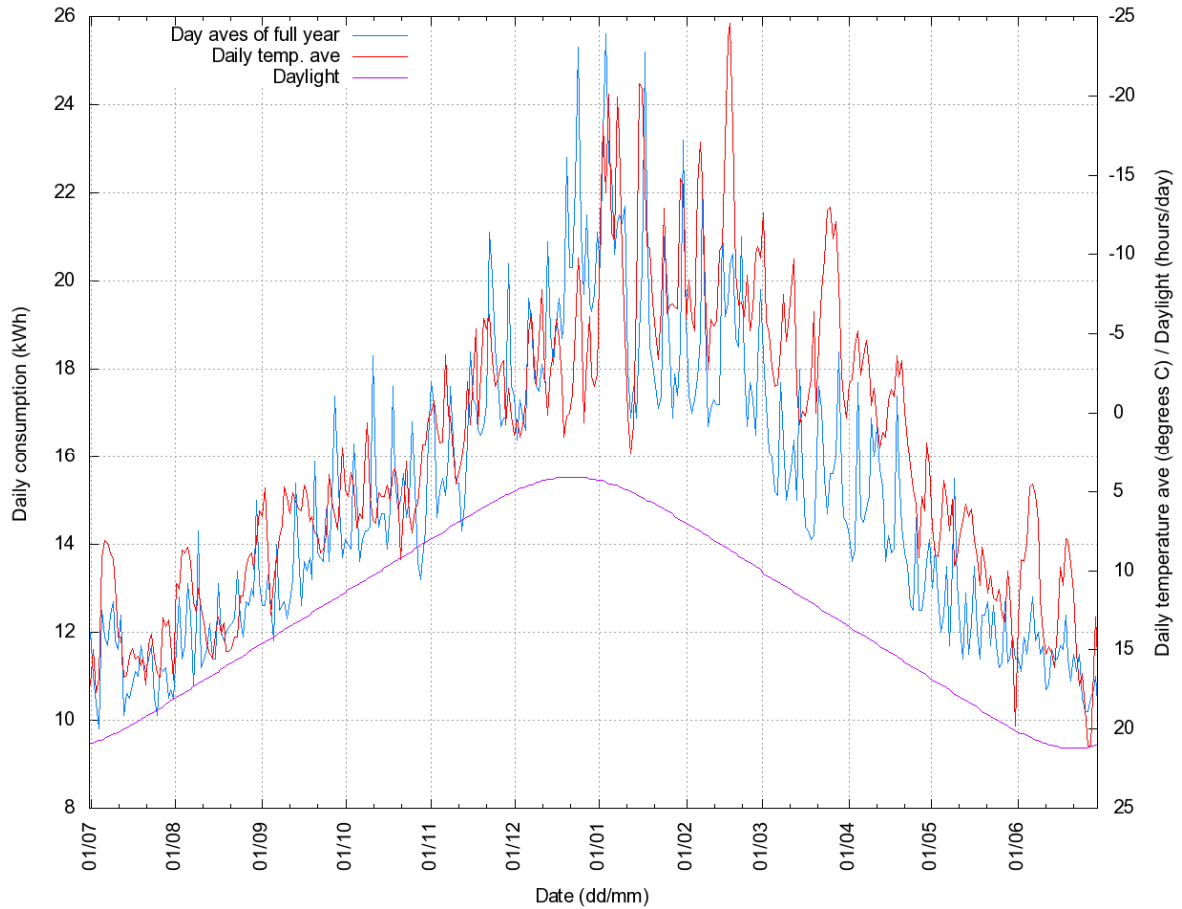
Kainuusta saadusta datasta erotettiin omaksi ryhmäkseen myös ilmalämpöpumppua käyttävät suorat sähkölämmittäjät. Näillä asiakkailla voi esiintyä yllättävän korkeita kulutuspiikkejä kylminä päivinä, jos lämpöpumppu ei ole mitoitetua kylmimpiä päiviä varten. 2008–2009 datassa tällaisia piikkejä ei kuitenkaan näkynyt. Lämpöpumppua käyttävien asiakkaiden sähkönkulutus

muuttui samalla tavalla kuin pelkkää suoraa sähkölämmitystä käyttävien ja päivärakenne oli samanlainen. 2008–2009 talvi oli kuitenkin lämmin, ja kylmän 2010 tammikuun sähkönkulutus oli hieman ennustettua korkeampi. Tätä käsitellään kappaleessa 7.4, mutta muuten lämpöpumppuasiakkaita ei käsitellä omana ryhmänä.



*Kuva 1: Keskimääräisen suoran sähkölämmittäjän päiväkulutus (kWh) vuoden aikana sinisellä. Punaisella lämpötilan päiväkeskiarvo, ja lilalla päivän pituus tunteina päivässä.*

Osalla suoraa sähkölämmitystä käyttävistä asiakkaista on ilmeisesti käytössä yösähkö esimerkiksi käyttöveden lämmitystä varten, koska keskiarvossa näkyy piikki yöllä yleensä kello 23. Sähkölämmittäjät, sekä varaava että suora, ovat päiväenergioiden vuosijakaumaa tarkasteltaessa keskenään melko samanlaisia., mutta päiväkulutuksen rakenteessa on suuria eroja eri tuntien välillä. Tämä käsitellään tarkemmin kappaleessa 6.1. Samoin öljy- ja kaukolämmittäjät ovat keskenään melko samanlaisia. Kuvassa 1 näemme keskimääräisen suoran sähkölämmittäjän ja kuvassa 2 keskimääräisen öljylämmittäjän vuosikulutuksen.



Kuva 2: Keskimääräisen öljylämmittäjän päiväkulutus (kWh) vuoden aikana sinisellä. Punaisella lämpötilan päiväkeskiarvo, ja lilalla päivän pituus tunteina päivässä.

Sähkölämmittäjä käyttää selkeästi enemmän sähköä, niin talvella kuin kesälläkin. Kulutus seuraa voimakkaasti lämpötilaa: hieman yllättäen myös kesällä. Öljylämmittäjän kulutus on kesällä vain hieman pienempi kuin suoran sähkölämmittäjän, mutta suuri ero näkyy talvella jolloin öljylämmittäjän kulutus on noin 2-3 kertaa pienempi. Öljylämmittäjänkin kulutus on kuitenkin talvella lähes kaksinkertainen kesään verrattuna. Kulutus näyttää osittain seuraavan lämpötilan muutosta, mutta koko vuoden mittakaavassa se näyttäisi seuraavan paremmin päivän pituuden muutosta. Tämä ei välttämättä johdu valaistuksen käytön muutoksista, vaan esimerkiksi televisiota saatetaan katsoa enemmän pimeinä päivinä. Päivänpituuden vaikutusta käsitellään tarkemmin kappaleessa 5.1

## 2.2 Asiakastyypin keskiarvojen analysointi

Tietyn asiakastyypin analyysiä käytetään ohjelman toiminnan testausta varten, mutta sitä voidaan käyttää myös antamaan arvio tietyn asiakastyypin keskimääräisen asiakkaan kulutuksesta. Näin se vastaa osittain samaa asiaa kuin viitteen [10] kappaleessa 3.9.2 esitelty kuormitusmallit. Kuormitusmalleja voidaan tällä ohjelmalla tehdä aina uusimmasta datasta.

Tässä ohjelmassa ei kuitenkaan käsitellä tietyn tyyppiseen asiakasryhmään kuuluvien asiakkaiden välisiä eroja. Jos asiakkaita on  $n$  kappaletta, määräytyy analyysissä käytettävän sarjan yhden tunnin kulutusarvo kaavasta

$$p_h = \frac{\sum_{i=1}^n p_{i,h}}{n}, \quad (\text{Kaava 2.1})$$

missä  $h$  on vuoden tietty yksittäinen tunti. Kaavan summatermillä on tietysti myös hajonta mutta sitä ei oteta tässä mallissa millään tavalla huomioon. Oli kyseessä sitten kolme asiakasta joiden kulutukset ovat tietyllä tunnilla  $h$  1, 2 ja 3 kWh tai kolme asiakasta joiden kulutukset ovat 2, 2 ja 2 kWh, on näille laskettu  $p_h$  sama.

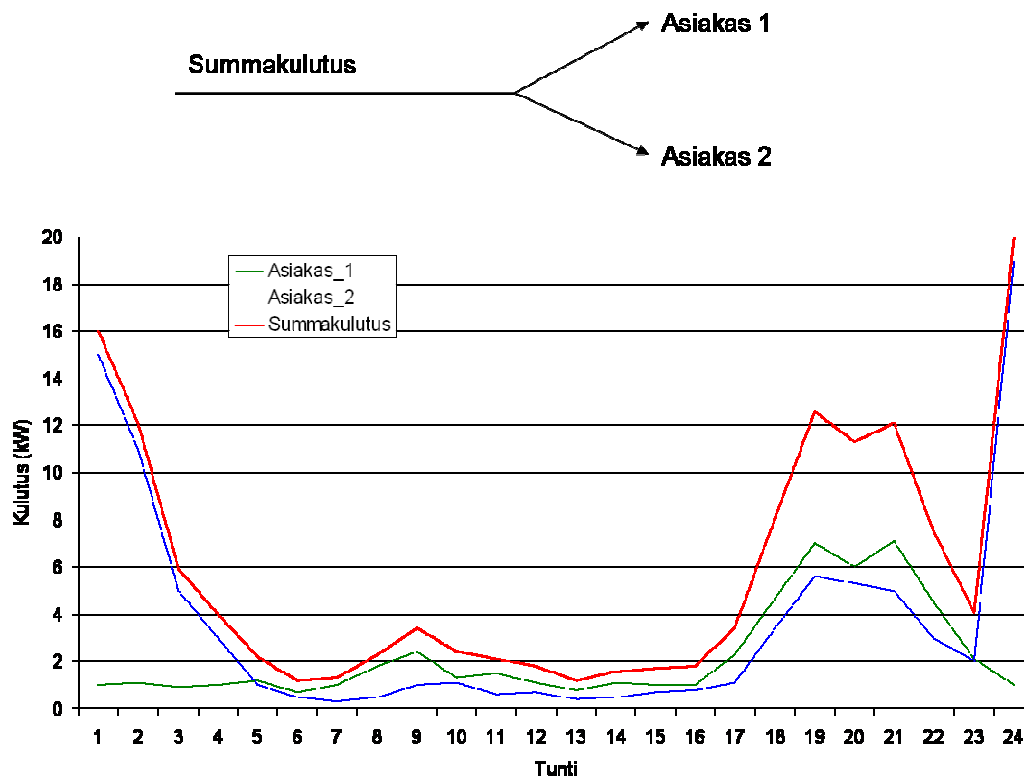
Ohjelmassa kuitenkin lasketaan varianssi tietyn tunnin keskiarvoa  $p_h$  laskettaessa. Tällöin keskiarvon laskentaa on mahdollista kehittää niin että asiakkaiden kulutuksen hajonta otetaan huomioon. Näin voitaisiin saada kuormitusmalleja vastaavat hajonnat sisältävät arviot eri asiakastyypeille. Tällä hetkellä ohjelman päätarkoitus on kuitenkin asiakkaiden summan analysointi jolloin eri asiakkaiden välistä hajontaa ei tarvitse tutkia, kuten kappaleessa 2.3 näemme.

## 2.3 Pienen asiakasryhmän summan analysointi

Tässä työssä on ison asiakasryhmän keskiarvon tutkimista tärkeämpänä pidetty pienen asiakasryhmän summan tutkimista. Tällöin ei ole tarvetta pohtia keskiarvoon liittyvän hajonnan huomioon ottamista. Summaa analysoitaessa yhden tunnin kulutusarvo on

$$p_h = \sum_{i=1}^n p_{i,h}. \quad (\text{Kaava 2.2})$$

Asiakasryhmän summaa tutkittaessa ei viitteessä [10] kappaleessa 3.9.2 käsitelty risteily vaikuta, kun kuvan 3 mukaisesti tutkitaan suoraan asiakkaiden summakulutusta. Kuitenkin sama ongelma risteilyn kanssa voi syntyä kun olemme analysoineet kulutuksen esimerkiksi jakelumuuntajatasolla ja haluaisimme summata nämä kulutukset yhteen. Ohjelma antaa ennusteen tietyn päivän tunnille ja myös hajonnan joka kertoo miten tietyn muuntajan kulutus vaihtelee valitun tyyppisten tuntien välillä. Tästä hajonnasta on poistettu lämpötilan ja päivänpituuden vaikutus ohjelman regressioanalyysin avulla, mutta jäljellä jäävä hajonta voi korreloida muiden muuntajien summien hajontojen kanssa. Esimerkiksi joulun aikaan useimpien asiakkaiden kulutus on suurimmillaan.



Kuva 3: Keksitty esimerkki kahden asiakkaan tuntikulutuksien risteilystä. Summaa käsiteltäessä ei tarvitse välittää yksittäisten asiakkaiden kulutushuipusta, vaan summan käsittely riittää.

Tässä työssä keskitytään kuitenkin ainoastaan yhden asiakasryhmän summan, esimerkiksi yhden muuntajan syöttämän alueen, ennustamiseen, jolloin risteilyä ei tarvitse huomioida; se näkyy suoraan summakulutuksessa. Kappaleessa 4.4 on kuitenkin käsitelty useiden asiakkaiden ja



asiakasryhmien summan hajonnan muodostumista teoreettisesti. Sen pohjalta ohjelmaa voidaan kehittää niin, että on mahdollista summata yksitellen käsiteltyjä jakelumuuntajien summia. Siinä myös nähdään miten eri asiakkaiden hajonnat näkyvät ryhmän summan hajonnassa, ja kuinka summaa käsiteltäessä pelkkä summan hajonnan käsittely riittää.

### 3 Vuoden päivien jakaminen ryhmiin

Tässä työssä analysoidaan lämpötilan ja päivän pituuden vaikutus sähkönkulutukseen olettaen lineaarinen suhde selittävän ja selitettävän välille. Lämpötilan vaikutus kulutukseen ei kuitenkaan ole lineaarinen koko vuoden aikana. Ainakin kesällä kun lämmitystä ei käytetä, ei lämpötila vaikuta kulutukseen.

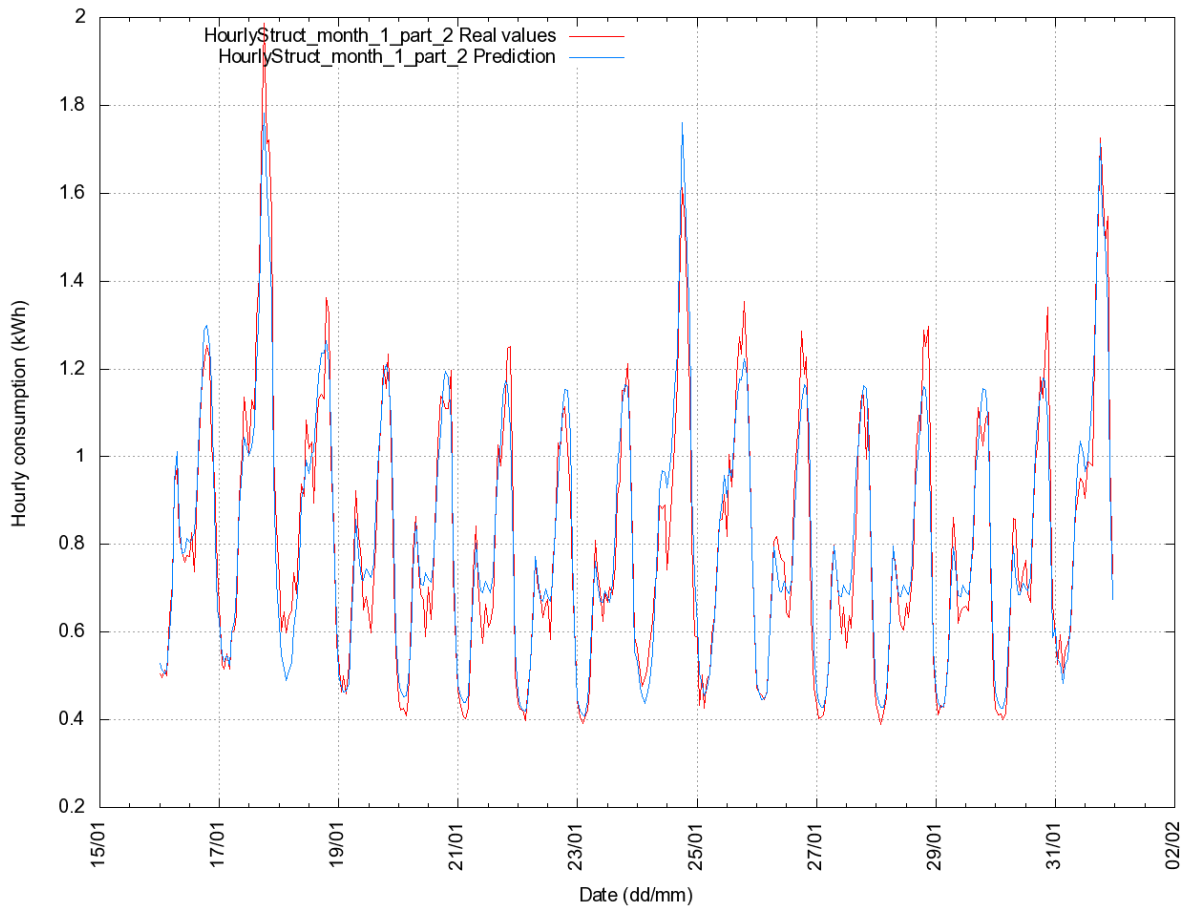
Lämpötila voitaisiin yrittää ottaa huomioon jotenkin muutenkin kuin lineaarista riippuvuutta käyttämällä, mutta tässä työssä on matematiikka yritetty pitää mahdollisimman yksinkertaisena. Epälineaarinen käyttäytyminen otetaan huomioon jakamalla vuosi osiin. Päivät jaetaan ryhmiin niin, että tietyn lämpötila-alueen sisään mahtuvat päivät kuuluvat samaan ryhmään. Tämän lisäksi eri päivätyypit jaetaan ryhmiin, koska arki-, aatto- ja pyhäpäivien rakenne on hyvin erilainen. Suurin mahdollinen ongelma on se, että johonkin ryhmään jää liian vähän päiviä. Tämä voi olla ongelma jos päivät jaetaan liian pieniin ryhmiin: esimerkiksi hyvin kylmiä aattopäiviä voi olla vuodessa hyvin vähän. Tällöin regressioanalyysin laskenta voi olla ongelmallista, ja hajonta voi jäädä suureksi.

#### 3.1 Jako arki-, aatto- ja pyhäpäiviin

Erikoistyössäni [5] jaoin päivät arki- ja lomapäiviin. Sekä lauantait että sunnuntait olivat samassa ryhmässä. Ohjelmassa on helppo jakaa päivät eri tavalla, ja suurimmat virheet näyttivät syntyvän lauantai ja sunnuntaipäiville jos niitä ei erotettu toisistaan. Päivien erilaisuus riippuu kuitenkin asiakastyypistä.

Tässä työssä on käytetty kotitalouksien mittausdataa ja tällöin kulutus on useimmiten kuvan 4 tyyppistä, eli lauantait erottuvat melko selkeästi muista viikonpäivistä. Toimistoilla ja pienillä yrityksillä ero olisi tietenkin vielä suurempi. Joissain kuvissa eri päivätyypit on erotettu englanninkielisten sanojen lyhenteillä: Workdys tarkoittaa arki-, Eves aatto- ja Holidys pyhäpäiviä. Sanat ovat englanniksi koska kuvien piirtäminen on ohjelman sisäänrakennettu ominaisuus, ja ohjelma on tehty kokonaan englannin kielellä. Päivätyyppien erottelu on esitetty kappaleessa 5.3. Ongelmana tarkemmassa erottelussa on kuitenkin se, että yhteen ryhmään voi jäädä hyvin vähän päiviä.

Tässä työssä päivät jaetaan oletuksena arki, -aatto ja lomapäiviin, mutta ohjelmassa voidaan joustavasti muuttaa tätä jakoa ja laskea oma malli vaikka kaikille viikonpäiville erikseen. Aattopäiviin kuuluvat lauantaiden lisäksi uudenvuodenaatto ja vappuaatto. Jouluaattonaatto on usein päiväenergialtaan aattopäivän kaltainen joten se on tässä luettu aattopäivien joukkoon. Jouluaatto on pyhäpäivä, mutta päivän kulutuksen summan kannalta yleensä lauantain kaltainen joten se on luettu tässä aattopäiviin. Pyhäpäiviin kuuluvat sunnuntain lisäksi uudenvuodenpäivä, loppiainen, itsenäisyyspäivä, vappu, joulupäivä, tapaninpäivä, juhannusaatto, juhannuspäivä, pyhäinpäivä, pitkäperjantai, helatorstai ja 2. pääsiäispäivä.



*Kuva 4: Keskimääräisen öljylämmittäjän tunneittainen kulutus ajalla 16.1.2009-31.1.2009: punaiselle alkuperäinen mittaustiedot ja sinisellä mallin antama ennuste. 17.1., 24.1. ja 31.1. olivat lauantapäiviä. Mallissa päivät oli jaettu arki-, aatto- ja pyhäpäiviin.*

Varsinkin joulunajan pyhien ja aattojen kulutus riippuu voimakkaasti siitä minkä tyyppisiä asiakkaita käsitellään. Tässä työssä on käsitelty vain kotitalousasiakkaita joten esimerkiksi jouluaattona kulutus on melko korkeaa. Toisaalta yritysten tapauksessa jouluaatto pitäisi luultavasti olla pyhäpäivien kanssa samassa ryhmässä. Joulunajan pyhät voidaan joko jakaa eri

päiväryhmiin erityyppisten asiakkaiden tapauksissa tai jättää mallin ulkopuolella jotta ne eivät sotke muiden päivien analysointia.

Kotitalousasiakkailla jouluaatto ja aaton aatto ovat päiväenergialtaan lauantain kaltaisia, mutta eivät päivän rakenteeltaan. Tämän takia ne sotkevat jonkin verran kylmien aatto- ja pyhäpäivien analyysiä. Eri tuntien kulutus heittelee muutenkin niin paljon, että joulunajan päivien käsittelyä aattopäivinä on pidetty tarpeeksi hyvänä ratkaisuna. Toinen vaihtoehto voisi olla joulun ajan käsitteleminen täysin erikseen. Päivärakenteita on käsitelty tarkemmin kappaleessa 6.1.

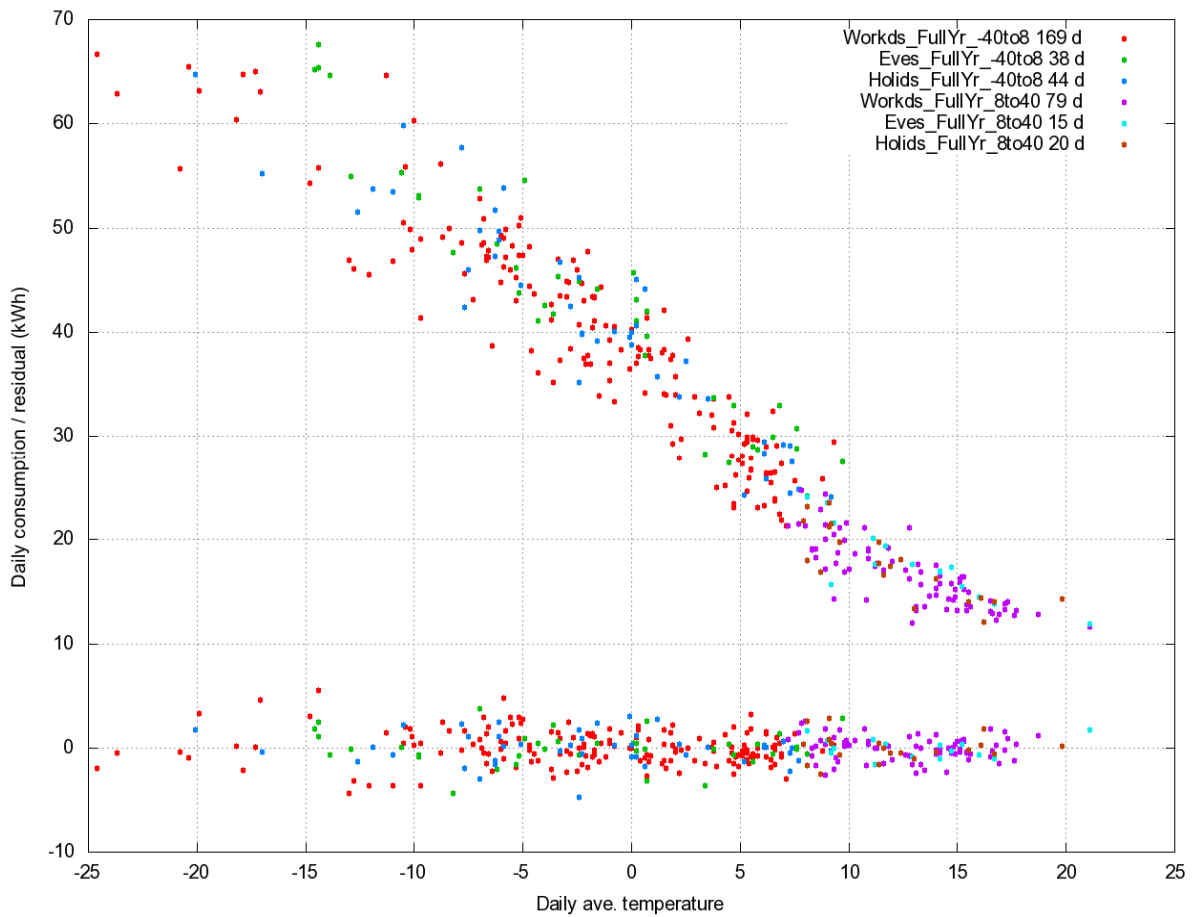
### 3.2 Vuoden päivien jakaminen lämpötila-alueisiin

Tässä työssä regressiolaskenta tehdään aina olettaen lineaarinen suhde selittävän ja selitettävän muuttujan välille. Ainakin lämpötila voi kuitenkin vaikuttaa epälineaarisesti. Kylmimpinä päivinä voi esimerkiksi lämpöpumppua käyttävän asiakkaan kulutus nousta lineaarista arvioita korkeammaksi koska pumpput saattavat toimia kylmällä ilmalla huomattavasti huonommalla hyötysuhteella. Hyvin kylmällä ilmalla, kun pumppu ei pysty lämmittämään huoneistoa haluttuun lämpötilaan, asiakas voi myös käyttää sähköpatteria lisälämmitykseen. Tämä epälineaarisuus on tässä työssä otettu huomioon jakamalla vuosi osiin ja käsittelemällä kukin osa olettamalla lineaarinen lämpötilariippuvuus sen sisällä.

Päivät voidaan jakaa ryhmiin monella eri tavalla. Ohjelmassa on mahdollista jakaa päivät vuoden aikojen mukaan neljään ryhmään tai lämpötilojen mukaan kahteen tai kolmeen ryhmään. Tässä työssä on päädytty jakamaan vuosi lämpötila-alueisiin, eikä vuodenaikojen mukaan. Tällöin saadaan sellaisia päiväryhmiä joissa myös valoisuus muuttuu ja kevään ja syksyn ero koetetaan selittää päivänpituuden avulla. Lämpötilojen mukaan jaettaessa määritellään piste tai pisteet jossa jako vaihtuu. Tästä on esimerkki kuvassa 5: jakopiste on +8 astetta. Tämän lisäksi päivät jaetaan yleensä arki-, aatto- ja pyhäpäiviin.

Jaettaessa päiviä pienempiin ryhmiin, vaikeutuu regressioanalyysi. Ongelmia voi aiheuttaa epätavallisten päivien suuri suhteellinen määrä. Tämän lisäksi tietynlaisissa jaoissa analyysi ei välttämättä pysty erottamaan päivänpituuden vaikutusta. Jaettaessa vuosi vuodenaikojen mukaan, ei talvella päivän pituuden vaikutusta välttämättä pysty kunnolla erottamaan. Tämän voi nähdä kuvasta 6: talvella valoisan ajan muutos on pientä johtuen sinityyppisestä käyrämuodosta. Regressioanalyysi etsii yhteyttä selittävien ja selitettävän välille ja jos selittävä muuttuja muuttuu

hyvin vähän, ei yhteyttä välttämättä löydetä. Tämän lisäksi joulun ja uuden vuoden epätavalliset kulutukset voivat sotkea talven käsittelyä.



Kuva 5: Keskimääräisen suoran sähkölämmittäjän päiväkulutus (kWh), kun x-akselin arvoiksi on valittu päivän lämpötilakeskiarvo, näkyy tässä ylempänä vinona joukkona pisteitä. Vuosi on jaettu kahteen lämpötila-alueeseen kohdasta +8 astetta, sekä kolmeen eri päivätyyppiin: esim. Holid\_FullYear\_-40to8 sisältää ne vuoden pyhäpäivät joiden lämpötila oli välillä -40..+8 astetta. Alempi joukko pisteitä esittää ennusteen tekemää virhettä eli residuaalia päiväkulutusten ennustamisessa eri lämpötiloissa.

Ohjelmaa testattaessa on kokeiltu kahteen ja kolmeen lämpötila-alueeseen jakamista. Kummassakin tapauksessa on lämpimät päivät eroteltu omaksi ryhmäkseen kohdasta +8 astetta koska kesällä lämpötila vaikuttaa kulutukseen eri lailla kuin silloin kun lämmitys on päällä. Kolmeen lämpötila-alueeseen jaettaessa päivät jaetaan vielä kohdasta -2 astetta. Kuvan 5 mukainen jako kahteen alueeseen laittaa kaikki alle +8 asteen päivät samaan ryhmään, joten mahdollista hyvin kylmän sään epälineaarista riippuvuutta ei saada otettua huomioon. Voimakas epälineaarinen käyttäytyminen on kuitenkin melko harvinaista ja se näkyy vasta niin kylminä

päivinä että sen käsitteleminen omana ryhmänä on vaikeaa. Liitteessä 1 näemme varaavan sähkölämmittäjän päiväenergioiden ennusteen kun vuosi on jaettu kolmeen lämpötila-alueeseen.

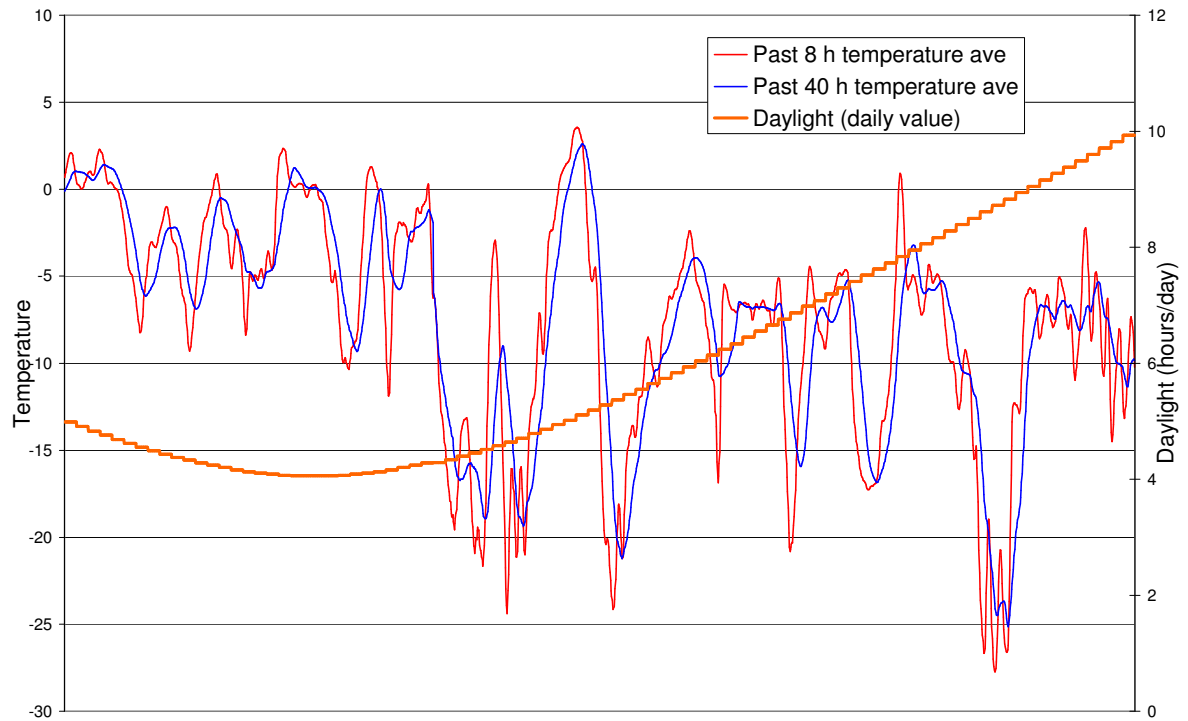
Käyttämällä jakoa kahteen lämpötila-alueeseen eli ottamalla kaikki ne päivät joina lämmitystä käytetään omaksi ryhmäkseen takaa, että kylmän sään analyysiä varten jää riittävä määrä päiviä. Jako kolmeen lämpötila-alueeseen tekee analyysin vaikeammaksi, ja sitä ei ole juurikaan käytetty tässä työssä näkyvissä kuvissa. Jos jakoa kolmeen lämpötila-alueeseen on käytetty, siitä on kerrottu kuvatekstissä. Kolmen lämpötilan-alueen käytön puolesta puhuu kuitenkin vuoden 2010 tammikuun testi varaavan sähkölämmityksen osalta (kts. kappale 7.3), joten ei ole täysin selvää kumpaa jakoa kannattaisi käyttää. Jako kahteen lämpötila-alueeseen toimii varmemmin, mutta kolmeen alueeseen jako antaa joissain tilanteissa selvästi paremman tuloksen.

Kuvan 5 alemmassa osassa näkyy suoran sähkölämmittäjän melko lineaarinen riippuvuus lämpötilan ja kulutuksen välillä. Kulutus on esitetty lämpötilan funktiona. Kesällä, kun lämmitystä ei käytetä, lämmitys ei kuitenkaan vaikuta kovin vahvasti kulutukseen. Kuvasta katsomalla voisi tälle asiakastyypille lämpötilarajan laittaa hieman +8 astetta korkeammallekin, mutta tässä on haluttu varmistaa että myös kesän ryhmään tulee kohtuullinen määrä päiviä regressioanalyysiä varten.

### 3.3 Kahden lämpötilamuuttujan käyttö

Erilaisten rakennusten lämmönvaraamiskykyjen huomioon ottamiseksi oli tarkoitus käyttää kahta selittävää lämpötila-arvoa, esimerkiksi 8 ja 40 tunnin lämpötilakeskiarvoa. Tarkoitus oli matemaattisen analyysin avulla pystyä painottamaan näitä selittäviä tekijöitä niin, että saadaan parhaimman selityksen antava lopputulos. Tällöin selittäviä tekijöitä olisi ollut yhteensä kolme kappaletta.

Kahden lämpötilamuuttujan käyttö aiheutti kuitenkin joissain tilanteissa matemaattisia ongelmia regressioanalyysissä. Kuvassa 6 nähdään miten samalla tavalla 8 ja 40 tunnin lämpötilakeskiarvot muuttuvat. Tällöin ongelmaksi voi nousta multikollinearisuus. Multikollinearisuus [6] tarkoittaa tilannetta, jossa regressioanalyysin selittävät muuttujat korreloivat voimakkaasti keskenään, eikä regressioanalyysin lopputuloksesta saada järkevää.



Kuva 6: Päivän pituus tunteina päivässä sekä kaksi eri lämpötilakeskiarvoa ajalla 1.12-28.2.2009. Lämpötila-arvot ovat tunneittaisia, mutta päivän pituus on tietynä päivänä kaikille vuorokauden tunneille sama koska se lasketaan kaavan 5.1 mukaisesti päivittäisenä arvona.

Kahden selittävän muuttujan välisen korrelaation ollessa suurta, ei regressioanalyysi välttämättä pysty erottamaan näiden vaikutusta toisistaan. Joskus matemaattinen analyysi antoi lyhyen ja pitkän ajan lämpötilavaikutukselle eri suunnan, eli kylmemmässä säässä kulutus olisikin pienempää. Tässä työssä analyysin onnistumisen varmuus on tärkeämpää kuin tarkkuus, joten kolmen muuttujan regressioanalyysistä luovuttiin. Multikollineaarisuus voi periaatteessa olla ongelma myös yhtä lämpötilamuuttujaa käytettäessä, koska tällöinkin selittäviä muuttujia on päivän pituuden takia kaksi, mutta tällaista ongelmaa ei ole tullut esiin.

### 3.4 Lämpötilan vaikutuksen hitaus

Lämpötila ei vaikuta kulutukseen saman tien, vaan vasta tietyn viiveen kuluttua. Tämä viive riippuu talon lämmönvaraamiskyvystä ja eristyksestä, ja se on eri asiakkailta erilainen. Tässä työssä yritettiin ensin kahden lämpötilan, lyhyen ja pitkän ajan keskiarvon käyttämistä regressioanalyysissä. Kolmen muuttujan regressioanalyysin ongelmien takia päädyttiin laskemaan regressio muutamalla eri tavalla käyttämällä päivänpituutta ja yhtä lämpötila-arvoa.

Työssä näkyvissä tuloksissa on laskettu aina viisi eri laskentaa, käyttämällä 8, 16, 24, 42 ja 40 tunnin lämpötilakeskiarvoja. Näin on yritetty ottaa huomioon lämpötilan vaikutuksen hitaus. On oletettu että ei kannata käyttää suoraan esimerkiksi 16 tuntia vanhaa tietyn tunnin lämpötila-arvoa vaan on parempi käyttää keskiarvoa, koska lämpötila tuskin vaikuttaa kulutukseen yhtäkkisenä piikkinä. Yleensä yöllä parhaiten selittää pitkän aikavälin keskiarvo ja varsinkin illalla lyhyen ajan keskiarvo. Jos tunneittaista lämpötiladataa ei ole saatavilla, voidaan käyttää päivän lämpötilakeskiarvoa. Tällöin analyysi tehdään saman päivän, edellisen päivän ja näiden keskiarvon avulla ja näistä valitaan paras sovitus.

Paras sovitus valitaan sen perusteella, mikä viive antaa suurimman Pearsonin korrelaatiokertoimen neliön (kaava 5.10). Regressiolaskentaa varten on kuitenkin määriteltä joitain vaatimuksia jotka sovituksen lopputuloksen pitää täyttää ennen kuin sitä edes verrataan muihin sovituksiin. Nämä vaatimukset on määriteltä kappaleessa 5.4.



## 4 Hajonnan tutkiminen

Hajonnan tutkiminen on ennustetta tehtäessä erittäin keskeistä. Tietyn hetken sähkönkulutuksen ennusteen lisäksi on pystyttävä arviomaan ennusteen luotettavuus. Normaalijakautuneelle hajonnalle on helppo määrittää luottamustaso [7]. Tärkeimmäksi kysymykseksi nouseekin voiko ennusteen residuaalin, eli eron mallin ennusteen ja mitatun datan välillä, olettaa käyttäytyvän normaalijakauman mukaisesti.

Analyysissä käytettävän datan yksittäiset tunneittaiset arvot on määriteltä joko kaavan 2.1 tai 2.2 mukaan riippuen siitä tutkitaanko asiakastyypin keskiarvoa vai ryhmän summaa. Ryhmän summaa käsiteltäessä hajonnan määritelmä on suoraviivainen, ja sillä tarkoitetaan tietyn päiväryhmän tietyn kellonajan residuaalien keskihajontaa. Residuaalit kuvaavat sitä osaa mitatusta sähkönkulutuksesta, jota ei voitu selittää lämpötilan ja päivänpituuden avulla. Jos vuosi on jaettu esimerkiksi kahteen lämpötila-alueeseen kohdasta +8 astetta ja kolmeen päivätyyppiin, määritellään esimerkiksi alle +8 asteen aattopäivien kello 13 sähkönkulutukselle sekä kulutusennuste että residuaalien keskihajonta, joka kuvaa ennusteen epävarmuutta.

Jos käytetään asiakkaiden keskiarvoja, hajonta voi periaatteessa tarkoittaa joko eri päivien välistä hajonta (kuten edellä esitettyssä asiakkaiden summakulutuksen tapauksessa) tai eri asiakkaiden välistä hajontaa. Ohjelmassa ei kuitenkaan huomioida millään tavalla eri asiakkaiden välistä hajontaa. Keskiarvojen käsittely on lähinnä testausta varten, ja silloin ryhmän asiakkaiden on oletettu olevan melko samanlaisia ja niitä on oletettu olevan melko paljon. Tällöin ei asiakkaiden välisten erojen käsittelyä ole pidetty kovin tärkeänä. Kappaleessa 4.4 on kuitenkin käsitelty teoreettisesti ryhmän asiakkaiden tai osaryhmien välistä hajontaa.

Pienen asiakasryhmän summan käsittely on ohjelman tärkein toiminnallisuus, ja summaa käsiteltäessä eri asiakkaiden välistä hajontaa ei tarvitse ottaa huomioon edes periaatteessa. Kaikki kuviin (esimerkiksi kuvat 15 ja 17) merkityt hajonnat tarkoittavat siis tietyn ryhmän eri päivien hajontaa, käsiteltynä jokaiselle 24 tunnille erikseen.

## 4.1 Normaalijakauman oletaminen

Normaalijakaumassa suurin osa arvoista osuu hyvin lähelle keskiarvoa. Tässä työssä arvioidaan jokaisella vuoden tunnille jokin sähkönkulutuksen arvo sekä ennusteen epävarmuus. Epävarmuus kuvataan residuaalien keskihajontana, ja sitä voidaan käyttää määrittäessä haluttu luottamustaso [7], mutta tällöin residuaalien on oletettu jakautuvan normaalijakauman mukaisesti.

Normaalijakaumaoletusta käytetään tässä työssä lähtökohtana mutta ohjelmaan on sisällytetty mahdollisuuksia analysoida kuinka lähellä normaalijakaumaa ollaan. Tällä hetkellä ohjelma ei anna varoitusta liian suuresta määrästä yllättävän korkeita arvoja, mutta tunneittaista kulutusennustetta analysoitaessa nähdään että niitä voi esiintyä normaalijakaumaan nähden liian paljon. Kappaleessa 4.3 analysoidaan tunneittaisia virheitä tarkemmin.

Ohjelmaan voidaan helposti lisätä toiminnallisuus joka varoittaa liian suuresta määrästä yllättävän korkeita kulutusarvoja. Tällöin voidaan ilmoittaa esimerkiksi että laskettu 95 % luottamusväli ei ole käyttökelpoinen koska korkeita piikkejä voi olla enemmän kuin niille jäävä 2,5 % osuus. Ohjelmassa on jo analyysityökalut tätä varten, mutta vielä pitää määrittää kuinka suuri määrä tietyn tason, esimerkiksi  $\mu + 3\sigma$ , ylittäviä arvoja sallitaan. Esimerkiksi yli  $\mu + 5\sigma$  arvoja ei edes kohtuullisesti normaalijakaumaa seuraavassa datasarjassa saisi olla, koska näitä ei pitäisi tunneittaisilla arvoilla sattua kertaakaan vuoden aikana [7].

## 4.2 Päiväenergioiden hajonta

Taulukossa 1 on esitetty neljän eri lämmitystyyppin päiväenergioiden residuaalin keskihajonta. Taulukossa tämä hajonta on kilowattitunteina kaikilla asiakkailla hieman korkeampi talvella, eli hajonta näyttäisi hieman kasvavan kulutuksen kasvaessa. Ero on kuitenkin niin pieni, että selittämättä jääneen osan voidaan olettaa pysyvän suurin piirtein vakiona vuoden eri osissa. Näin asian pitäisikin olla, sillä lämpötilan ja päivänpituuden huomioon ottamisen jälkeen selittämättä jäävän osan ei kuuluisi kasvaa merkittävästi lämpötilan laskiessa ja päivän lyhentyessä vaan näistä aiheutuva kulutuksen kasvu tulisi pystyä selittämään.

Tässä työssä tehdään oletus, että residuaalien suuruus ei merkittävästi muutu lämpötilan tai päivänpituuden muutoksen vaikutuksesta käsiteltävässä päiväryhmässä. Tämä residuaalien

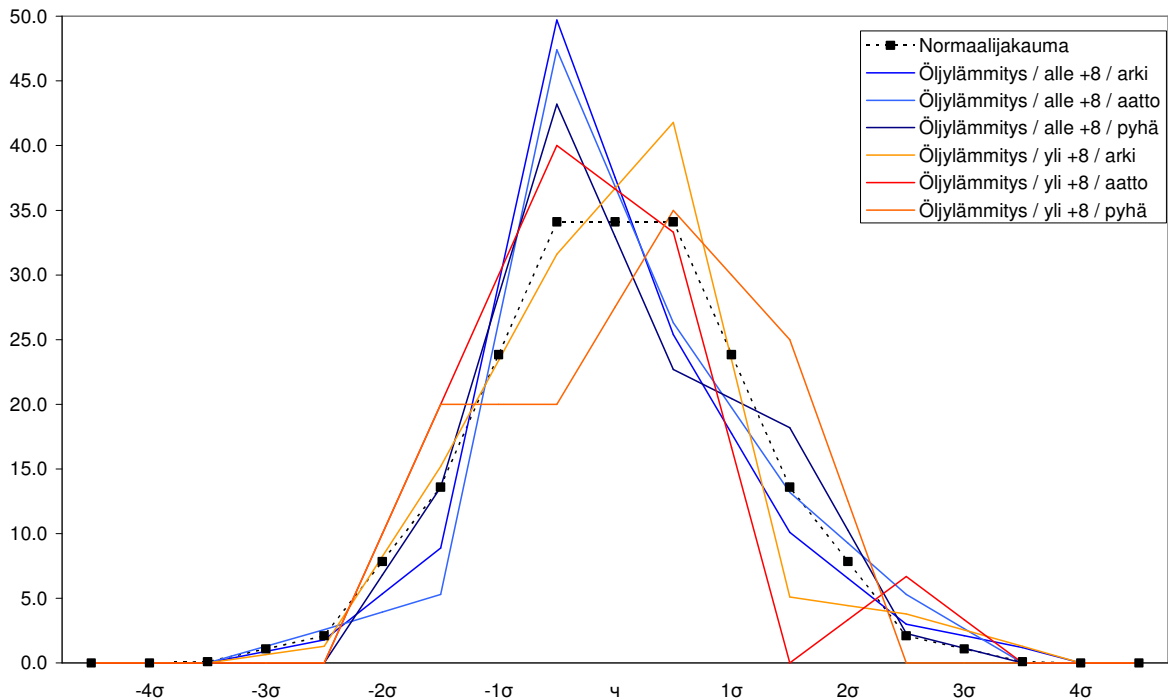
riippumattomuus selitettävistä muuttujista on usein käytetty oletus lineaarista regressioanalyysiä tehtäessä. Kuvien 5 ja 12 mukaan tämä ei kuitenkaan täysin pidä paikkaansa. Päivänpituuden muuttuessa pimeimpien päivien korkeammat kulutukset selittää joulun aika, mutta kylmien päivien korkeammat kulutukset ovat yllättäviä. Normaalijakauman mukaisesti määriteltä luottamustaso voi näin olla liian matala. Tämä voi olla vielä pahempi tilanne lämpöpumppua käyttävillä asiakkailla. Tätä tarkastellaan tarkemmin kappaleessa 7.4.

Kaukolämpö -40..+8		Öljylämmitys -40..+8	
Arki	9.80 ± 0.45 (4.6 %)	Arki	16.0 ± 0.83 (5.2 %)
Aatto	11.9 ± 0.47 (4.0%)	Aatto	19.1 ± 0.99 (5.2 %)
Pyhä	10.4 ± 0.39 (3.8 %)	Pyhä	17.2 ± 0.67 (3.9 %)
Kaukolämpö +8..+40		Öljylämmitys +8..+40	
Arki	7.60 ± 0.39 (5.1 %)	Arki	11.6 ± 0.50 (4.3 %)
Aatto	8.30 ± 0.27 (3.3 %)	Aatto	12.7 ± 0.40 (3.2 %)
Pyhä	7.51 ± 0.33 (4.3 %)	Pyhä	11.7 ± 0.54 (4.6 %)
Suora sähkölämmitys -40..+8		Varaava sähkölämmitys -40..+8	
Arki	39.0 ± 1.60 (4.2%)	Arki	51.0 ± 2.70 (5.2 %)
Aatto	43.0 ± 1.65 (3.8%)	Aatto	57.5 ± 2.64 (4.6 %)
Pyhä	41.2 ± 1.55 (3.8 %)	Pyhä	54.4 ± 2.64 (4.9 %)
Suora sähkölämmitys +8..+40		Varaava sähkölämmitys +8..+40	
Arki	16.7 ± 1.13 (6.8 %)	Arki	22.4 ± 2.18 (9.7 %)
Aatto	17.8 ± 0.94 (5.3 %)	Aatto	23.8 ± 2.71 (11.4 %)
Pyhä	17.7 ± 1.29 (7.3 %)	Pyhä	23.0 ± 2.21 (9.6 %)

*Taulukko 1: Tässä taulukossa näemme testattujen asiakastyypin päiväkulutuksen keskiarvon kyseisessä päiväryhmässä (kWh), residuaalien keskihajonnan (kWh) ja residuaalien keskihajonnan prosentteina suhteessa päiväkulutuksen keskiarvoon kyseisessä päiväryhmässä. Vuoden päivät oli jaettu kolmeen päivätyyppiin ja kahteen lämpötila-alueeseen kodasta +8 astetta, eli oli käytetty perusjakoa kuuteen päiväryhmään.*

Residuaalien keskihajonta lasketaan jokaiselle päiväryhmälle erikseen, ja haluttu luottamustaso määritellään kullekin ryhmälle olettamalla residuaaleille normaalijakauma ja riippumattomuus selittävistä muuttujista. Jos residuaalien kasvu lämpötilan kylmentyessä on ongelma, voidaan esimerkiksi käyttää jakoa kolmeen lämpötila-alueeseen, tai niiden hajonnan tutkiminen voidaan jakaa useampaan osaan kuin itse regressioanalyysin teko. Kuvan 5 alaosassa näkyvä suoran sähkölämmittäjän residuaalien kasvu lämpötilan kylmetessä on tässä työssä tutkituista lämmitystyypeistä pahin eikä sekään ole kohtuuton, joten on päätetty käyttää perusoletusta residuaalien riippumattomuudesta selittävistä muuttujista.

Kuvassa 7 on esitetty keskimääräisen öljylämmittäjän päiväenergioiden residuaalien hajonta verrattuna normaalijakaumaoletukseen. Tämä on tutkituista asiakastyypeistä vähiten normaalijakauman kaltainen. Yllättävän suuria arvoja on hieman liikaa, mutta ei kohtuuttomasti.  $\pm 4\sigma$  tai sitä suurempia arvoja ei ole ollenkaan. Päiväenergioiden kohdalla voidaan melko luottavaisin mielin käyttää normaalijakauman mukaisia luottamusatasoja [7].

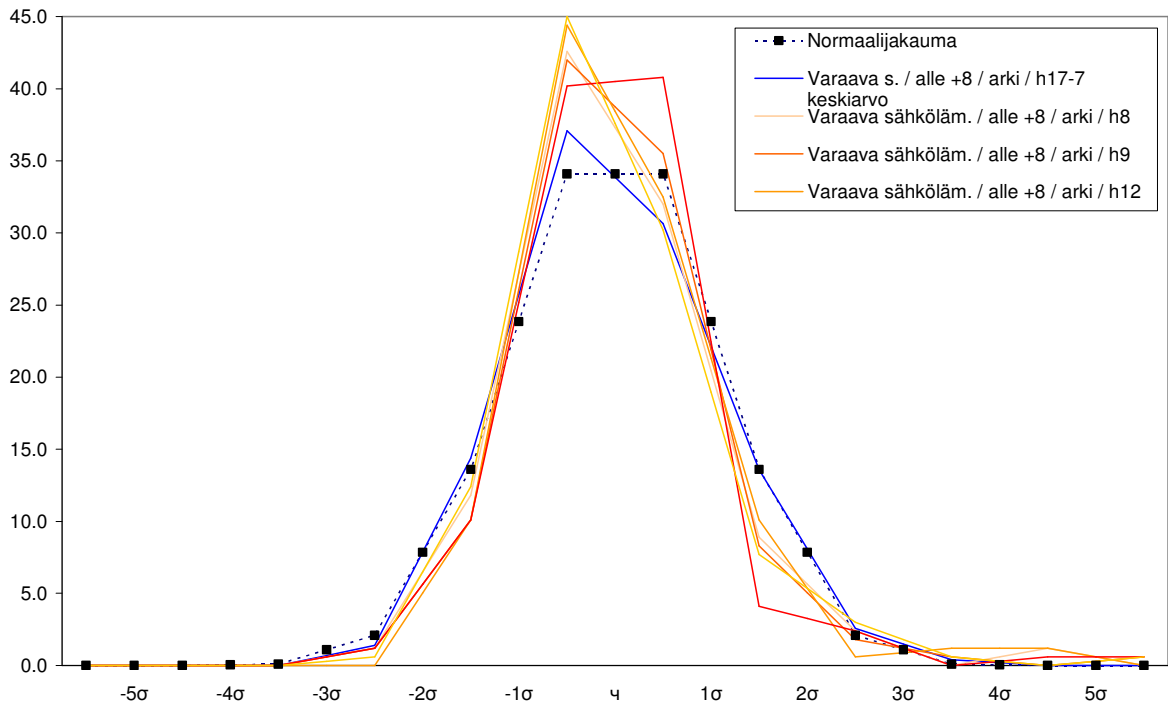


Kuva 7: Keskimääräisen öljylämmittäjän eri päiväryhmien päiväenergioiden residuaalien hajonta suhteessa normaalijakaumaan. Päiväryhmät on jaettu kuuteen perusryhmään. Y-akselilta voidaan lukea prosenttiosuus päivistä joiden residuaali osuu tietyn määrän hajontoja päähän keskiarvosta. Kuva on muodostettu jakamalla päivät yhden hajonnan tarkkuudella eri alueisiin, joten varsinkin kuvaajan keskiosa on melko epätarkka koska päivät voivat osua joko kohtaan  $\pm\sigma$  tai  $\pm 4\sigma$  mutta sen tarkempaa erottelua näiden välillä ei ole tehty. Katkoviivalla on esitetty referenssiksi normaalijakauma [7].

### 4.3 Tunneittainen hajonta

Ohjelmassa on toiminnallisuus joka tulostaa näkymän siitä mitä mallin eri tuntien residuaalien hajonnat ovat eri päiväryhmissä. Tämä tulostetaan tekstitiedostona, mutta tässä se on esitetty Excelin avulla kuvana. Esimerkiksi on otettu testatuista asiakasryhmistä vähiten normaalijakaumaa muistuttava tulos, joka oli varaavien sähkölämmittäjien keskiarvon arkipäivien ennusta lämpötila-alueella -40..+8 astetta: kuvassa 8 näemme ongelmallisimpien tuntien residuaalien hajontakuvion verrattuna normaalijakaumaan.

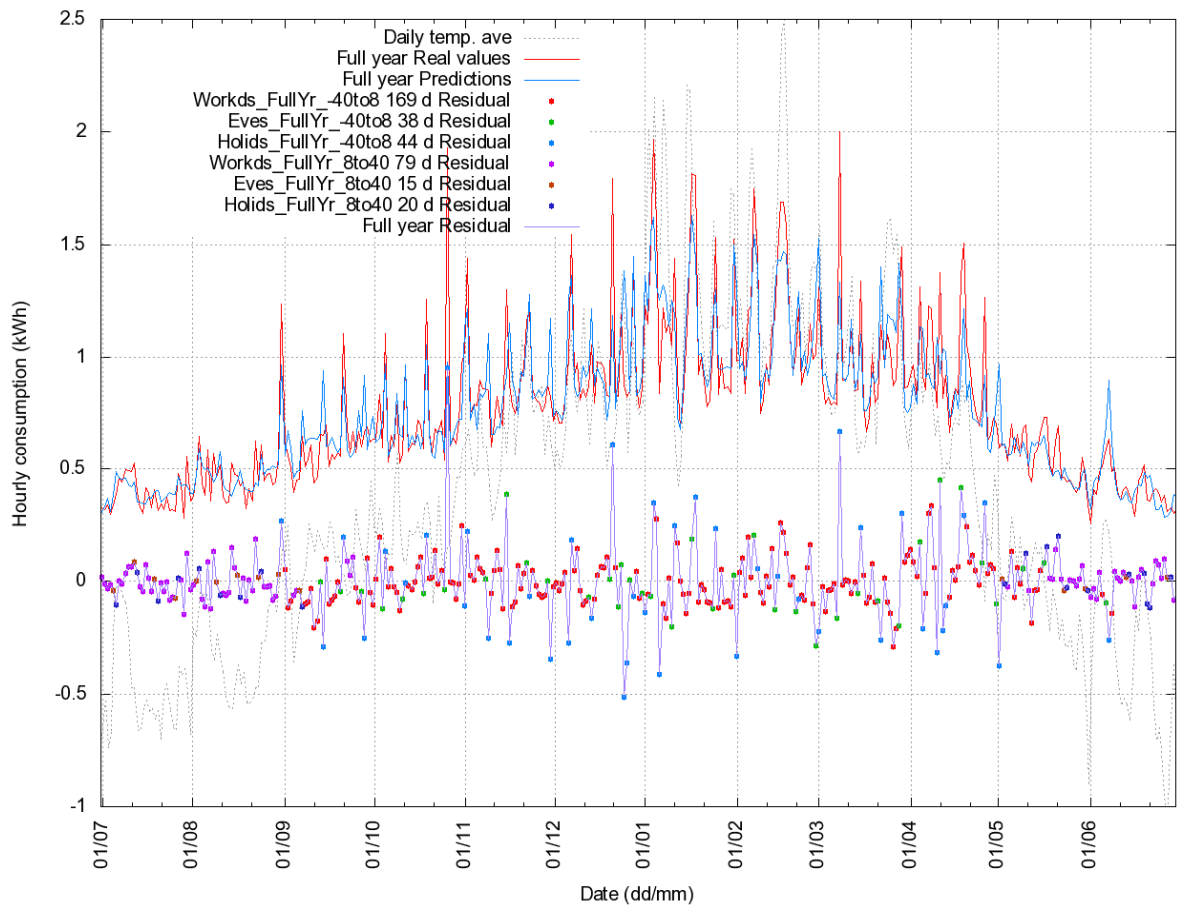
Kuvaan 8 on merkitty vertailun vuoksi muiden kuin kello 8-16 tuntien hajontarakenteen keskiarvo, ja huonoiten normaalijakaumaa seuraavat tunnit. Yllättävän suuret residuaalien arvot, eli yli neljä kertaa keskihajontaa suuremmat erot keskiarvosta, osuvat kaikki kello 8 ja 16 välille. Kaikkein suurimmat erot osuvat päivälle 2.1.2009. Kalenterista huomataan, että 2.1. oli arkipäivä, mutta se oli kahden vapaapäivän välissä eli 1.1. oli torstai joten 3.1. oli lauantai. Tästä voidaan päätellä että moni oli ottanut 2.1. torstain vapaaksi ja se näkyy keskimääräisten arkipäivien seassa vapaapäivänä, jolloin kello 8-16 välinen kulutus on yllättävän suurta. On kuitenkin huomattava että klo. 8-16 kulutus on varaavalla sähkölämmittäjällä melko pientä, kuten näemme kuvasta 17, joten silloin esiintyvät yllättävät kulutuspiikit eivät ole kovin haitallisia verkon mitoituksen kannalta. Yleensä korkean kulutuksen aika taas seuraa paremmin normaalijakaumaa, varsinkin sähköä lämmitykseen käyttävillä asiakkailla.



Kuva 8: Keskimääräisen varaavan sähkölämmittäjän tunneittainen residuaalien hajonta suhteessa normaalijakaumaan kylminä (alle +8 astetta) arkipäivinä. Sininen käyrä on tuntien klo. 0-7 sekä 17-23 keskiarvo ja muut käyrät ovat yksittäisiä tunteja väliltä klo. 8-16. Y-akselilta voidaan lukea prosenttiosuus tunneista jotka osuvat tietyn määrän hajontoja päähän keskiarvosta. Katkoviivalla on esitettyä normaalijakauma referenssiksi [7]

Muita yhtä suuria eroavaisuuksia normaalijakaumasta ei testeissä löytynyt, mutta lähes kaikilla lämmitystyypeillä oli hieman täydellistä normaalijakaumaa enemmän yllättävän suuria kulutusarvoja, joita ei pystytty selittämään lämpötilan tai päivänpituuden avulla. Tehtyjen testien perusteella tunneittainen kulutusennusteen ja mittauksien ero seuraa normaalijakaumaa, mutta

siihen sisältyy hieman liikaa yllättävän korkeita piikkejä. Tämän takia esimerkiksi normaalijakauman mukaan tehty arvio siitä että noin 95 % arvoista osuu alueelle  $\mu \pm 2\sigma$  [7], ei välttämättä pidä paikkaansa, vaan täytyy varautua suurempaan määrään yllättävän suuria arvoja. Ohjelmaan voidaan sisällyttää automaattinen varoitus jos jollekin tunnille sattuu esimerkiksi yli  $5\sigma$  virhe. Kuten edellä esitetystä 2.1.2009 tapauksesta nähdään, voi hyvin pieni eroavaisuus aiheuttaa piikkejä residuaalien arvoissa. On kuitenkin tärkeämpää tutkia huippukulutustuntien residuaaleja, kuten on tehty kappaleissa 6.2 ja 6.3.



*Kuva 9: Esimerkki hakevasti ennustettavasta tunnista: keskimääräisen varaavan sähkölämmittäjän kulutus vuoden eri päivinä kello 7. Punaisella mitattu ja sinisellä ennustettu tunneittainen kulutus, kun vuosi oli jaettu kahteen lämpötila-alueeseen ja kolmeen päivätyyppiin. Kuvan alaosassa on kuvattuna ennusteen tekemä virhe eli residuaali tunnilla numero 7 vuoden eri päivinä.*

Tietyn kellonajan ennusteen hyvyys on vuorokauden eri tunneilla hyvin erilainen. Sähköä lämmitykseen käytävillä asiakkailla yön tunnit pystytään yleensä ennustamaan hyvin, mutta päivällä ihmisten ollessa kotona kulutuksen ennustaminen on vaikeampaa. Kuvassa 9 on esitetty varaavan sähkölämmittäjän kello 7 esimerkkinä hakevasti ennustettavasta tunnista. Kuvassa

näkyä selkeä eroa pyhäpäivien ja muiden päivien välillä. Tämä ei ole kuitenkaan ole sinänsä ongelma, sillä eri päivätyypeille lasketaan oma residuaalien keskihajonta. Ylipäätään tuntien hajonta on selkeästi suurempaa ja vähemmän normaalijakauman kaltaista kuin päiväenergioiden hajonta. Kaikkien tuntien hajonta ei kuitenkaan ole yhtä vakavaa, vaan kannattaa keskittyä korkeiden kulutustuntien ennusteen hyvyyden tutkimiseen.

#### 4.4 Usean asiakkaan summan varianssi

Tähän työhön liittyvässä ohjelmassa ei käsitellä millään tavalla käsiteltävän asiakasryhmän eri asiakkaiden välistä hajontaa. On kuitenkin mahdollista, että tulevaisuudessa ohjelmaan halutaan lisätä mahdollisuus analysoida myös eri asiakkaiden välisiä eroja. Seuraavassa on esitelty tähän mahdollisesti sopivia teoreettisia menetelmiä.

Liitteessä [13] on käsitelty usean asiakkaan summan hajontaa. Summan hajonta voidaan laskea kaavasta [14]

$$\sigma_{Sum}^2 = Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \sum_{j=1}^n Cov(X_i, X_j) = \sum_{i=1}^n \sum_{j=1}^n \sigma_i \sigma_j \rho_{ij}, \quad (Kaava 4.1)$$

missä  $n$  on summattavien asiakkaiden määrä,  $\sigma_i$  on asiakkaan  $i$  hajonta ja  $\rho_{ij}$  on asiakkaiden  $i$  ja  $j$  välinen Pearsonin korrelaatiokerroin mikä voi saada arvoja väliltä  $-1 \dots 1$ .

Kun kaavan 4.1 mukaan edetään  $i$ :tä ja  $j$ :tä kasvattamalla, tulee vastaan pareja joissa  $i = j$ . Näissä summaustilanteissa voidaan käyttää kaavaa  $Cov(X, X) = Var(X)$  [15]. Näin kaava 4.1 saadaan muotoon

$$\sigma_{Sum}^2 = \sum_{i=1}^n Var(X_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n Cov(X_i, X_j) = \sum_{i=1}^n \sigma_i^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \sigma_i \sigma_j \rho_{ij}, \quad (Kaava 4.2)$$

koska  $\rho_{ii}$  on 1. Tästä nähdään että summan varianssi koostuu kahdesta osasta. Asiakkaiden ollessa täysin riippumattomia kaikki kovarianssit ovat nolliä, jolloin kaavan 4.2 jälkimmäinen summattava osuus tulee nolllaksi ja summan varianssi on asiakkaiden varianssien summa.

Jos käytetään varianssin ja kovarianssin keskiarvoa, voidaan kaava 4.2 kirjoittaa muotoon

$$\sigma_{Sum}^2 = n\bar{\sigma}^2 + n(n-1)\bar{\Phi}, \quad (\text{Kaava 4.3})$$

missä  $\bar{\Phi}$  on ryhmän asiakkaiden välisten kovarianssien keskiarvo. Jos edellä esitettyjä ryhmiä on  $k$  kappaletta, tulee näiden summan varianssiksi sijoittamalla kaava 4.3 kaavaan 4.4

$$\sigma_{Sum}^2 = \sum_{i=1}^k n_i \bar{\sigma}_i^2 + \sum_{i=1}^k n_i (n_i - 1) \bar{\Phi}_{ii} + \sum_{i=1}^k \sum_{\substack{j=1 \\ j \neq i}}^k n_i n_j \bar{\Phi}_{ij}, \quad (\text{Kaava 4.4})$$

missä  $n_i$  on ryhmän  $i$  asiakkaiden lukumäärä,  $\bar{\sigma}_i$  on ryhmän  $i$  asiakkaiden keskimääräinen hajonta,  $\bar{\Phi}_{ii}$  on ryhmän  $i$  asiakkaiden välinen keskimääräinen kovarianssi ja  $\bar{\Phi}_{ij}$  on ryhmien  $i$  ja  $j$  asiakkaiden välinen keskimääräinen kovarianssi.  $\bar{\Phi}_{ij}$  voidaan määrittää tarkkuutta menettämättä laskemalla ryhmien  $i$  ja  $j$  keskimääräisten asiakkaiden välinen kovarianssi. Tämä vähentää tarvittavien kovarianssilaskujen määrää.

Kaava 4.4 on sama kuin viitteessä [13] esitetty kaava numero 5, mutta tässä esitetyssä versiossa kovarianssi on laskettu kaavalla  $\Phi_{ij} = \sigma_i \sigma_j \rho_{ij}$ . Laskemalla suoraan kovariansseilla saadaan parempi tulos, koska viitteen [13] mukaisesti summaa laskettaessa hajontojen ja korrelaatioiden tulojen summa ei anna tarkasti oikeaa tulosta.

Kaavan 4.4 kaksi ensimmäistä summatermiä sisältävät kunkin ryhmän sisäiset kovarianssit ja viimeinen termi sisältää eri ryhmien väliset kovarianssit. Jos kukin ryhmä käsitellään tässä työssä esitetyssä ohjelmassa erikseen, tulee kaksi ensimmäistä termiä käsiteltyä automaattisesti, mutta ryhmien väliset kovarianssit tulee käsitellä erikseen. Ryhmien välisten kovarianssien laskemiseksi ei tarvitse vertailla yksittäisiä asiakkaita toisten ryhmien asiakkaisiin, vaan ryhmien keskiarvojen vertailu riittää eli

$$\bar{\Phi}_{ij} = \frac{1}{n_i n_j} \sum_i \sum_j Cov(X_i, X_j) = Cov(\bar{X}_i, \bar{X}_j), \quad (\text{Kaava 4.5})$$

missä  $i$  kulkee ryhmään  $i$  kuuluvien asiakkaiden läpi ja  $j$  ryhmään  $j$  kuuluvien asiakkaiden läpi niin että  $i$  on aina eri suuri kuin  $j$  eli kukin asiakas kuuluu vain yhteen ryhmään ja



$\overline{X}_i = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$  niin että käydään läpi kaikki ryhmään  $i$  kuuluvat  $n$  asiakasta.

Näin saadaan ryhmän keskiarvoasiakas.

Kun ryhmät käsitellään yksitellen, saadaan jokaisen ryhmän keskiarvoasiakkaalle hajonta ja varianssi. Tämän lisäksi voidaan laskea eri ryhmien välinen kovarianssi. Jos nyt halutaan laskea usean ryhmän summan varianssi, voidaan käyttää kaavasta 4.4 johdettua kaavaa

$$\sigma_{Sum}^2 = \sum_{i=1}^k n_i^2 Var(\overline{X}_i) + \sum_{i=1}^k \sum_{\substack{j=1 \\ j \neq i}}^k n_i n_j \overline{\Phi}_{ij}, \quad (Kaava 4.6)$$

missä  $\overline{\Phi}_{ij}$  lasketaan kaavan 4.5 mukaisesti,  $\overline{X}_i$  on ryhmän  $i$  keskiarvoasiakas ja  $n$  on ryhmään kuuluvien asiakkaiden lukumäärän. Tästä kaavasta voidaan erottaa eri ryhmien sisäisten varianssien ja toisaalta eri ryhmien välisten kovarianssien vaikutukset summan varianssiin.

Jos ryhmiä on vain yksi, saa kaava 4.6 muodon

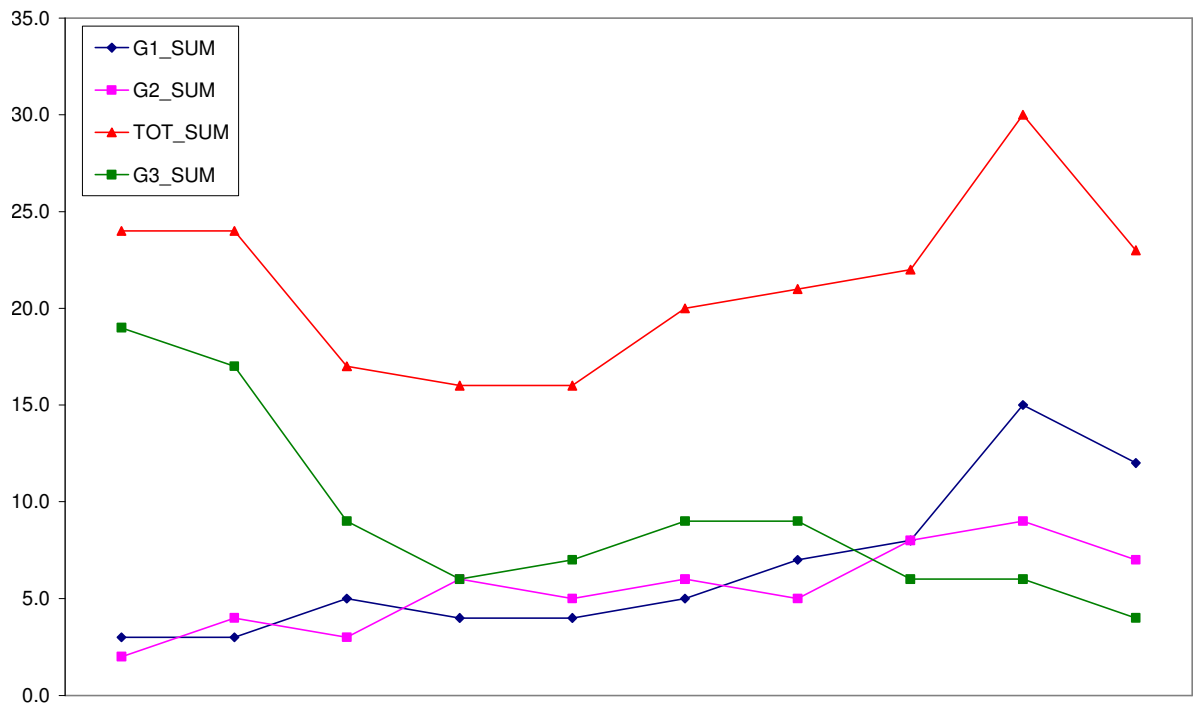
$$\sigma_{Sum}^2 = n^2 Var(\overline{X}) = Var(X_{Sum}), \quad (Kaava 4.7)$$

eli lopulta, jos kaikki asiakkaat summataan ensin yhteen, tarvitsee vain laskea tämän summan varianssi ja se pitää sisällään kaikkien yksittäisten asiakkaiden varianssit.

Kaavan 4.6 mukaisesti summan hajonta koostuu osista. Ryhmän sisäinen varianssi kasvattaa aina summan varianssia, mutta ryhmien välinen kovarianssi voi joko lisätä tai vähentää summan varianssia. Kuvassa 10 ja taulukossa 2 on esitetty keksitty esimerkki siitä, miten summan varianssi muodostuu. Koska yksittäisten ryhmien varianssi on yleensä selvillä, on summan varianssia laskettaessa laskettava kaikkien eri ryhmien välinen kovarianssi kaavan 4.6 jälkimmäisen summatermin mukaisesti. Vaihtoehtoisesti voidaan summata ryhmät yhteen ja laskea tälle kaikkien asiakkaiden ja kaikkien ryhmien yhteiselle summalle suoraan varianssi kaavan 4.7 mukaisesti.

Jos ryhmät lasketaan erikseen, saadaan tieto siitä miten eri ryhmät ja eri ryhmien väliset kovarianssit vaikuttavat koko summan varianssiin. Nämä komponentit on kuvan 10 mukaisessa

tilanteessa merkitty taulukkoon 2. Nähdään että summan varianssi ja sitä kautta hajonta on pienempi kuin yksittäisen ryhmän maksimihajonta (tässä ryhmä numero 3). Muut ryhmät tavallaan tasoittavat ryhmän numero 3 hajontaa, mikä näkyy negatiivisena ryhmien 1 ja 3 sekä 2 ja 3 välisenä kovarianssina.



Kuva 10: Kolmen kuvitteellisen asiakasryhmän G1, G2 ja G3 summat sekä niiden kokonaissumma kymmenellä eri ajankohdalla. Tämän tilanteen varianssi näkyy taulukossa 2.

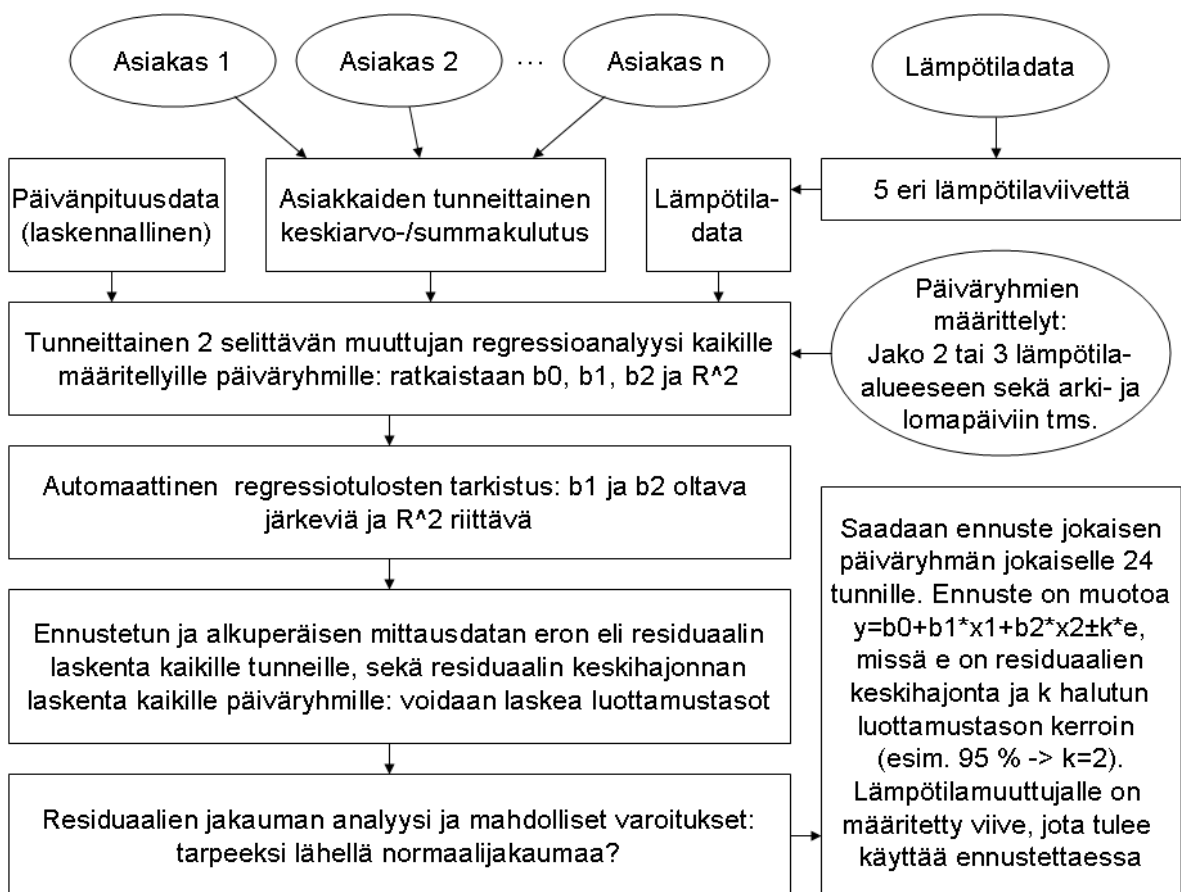
Kolmen ryhmän summa:					
Keskiarvo =	21.3	Keskihajonta =	4.1	Varianssi =	17.0
Summan varianssin komponentit:					
G1 oma vaikutus =	14.6	G2 oma vaikutus =	4.3	G3 oma vaikutus =	22.0
G1-G2 vaikutus =	12.4	G1-G3 vaikutus =	-21.6	G2-G3 vaikutus =	-14.6

Taulukko 2: Kolmen kuvitteellisen ryhmän G1, G2 ja G3 summan varianssi muodostuu kuuden komponentin summasta kaavan 4.6 mukaisesti: G1 oma vaikutus tarkoittaa ryhmän 1 varianssia ja G1-G2 vaikutus ryhmien 1 ja 2 välistä kovarianssia kaksinkertaisena jne.

Tässä kappaleessa esiteltiin mahdollisuus laskea usean ryhmän summan varianssi käsittelemällä ensin osaryhmiä. Tätä voi myös käyttää käsittelemällä yksittäisiä asiakkaita ja näiden välisen lasketun korrelaation avulla koota koko summan varianssi ja sitä kautta hajonta. Tämän työn yhteydessä tehdyssä ohjelmassa kuitenkin lasketaan aina suoraan koko summan varianssi kaavan 4.7 mukaisesti.

## 5 Ennusteen muodostaminen

Tässä työssä on kirjallisen osan ohella tehty C++ ohjelma tunneittaista sähkökuorman ennustamista varten. Sen toteutuksessa on käytetty työssä esitettyjä periaatteita. Regressioanalyysi tehdään käyttämällä kahta selittävää muuttujaa: lämpötilaa ja päivänpituutta. Lämpötilan vaikutuksen hitauden huomioon ottamiseksi tehdään useita regressioanalyysyjä käyttämällä eli viivettä kappaleessa 3.4 esitettyjen periaatteiden mukaisesti. Päivänpituuden vaikutuksen analyysissä ei käytetä mitään viivettä. Ohjelman toiminnan periaate nähdään kuvassa 11.



Kuva 11: Periaatekuva ohjelman toiminnasta. Ympyrämuotoiset objektit ovat käyttäjän antamaa dataa tai määrittelyitä, ja neliömuotoiset taas ohjelman sisäistä toimintaa.

Kuvan 11 mukaisesti tiettyyn ryhmään kuuluvien, esimerkiksi yhden jakelumuuntajan syöttämien, asiakkaiden mittausdata kootaan yhdeksi aikasarjaksi. Kappaleen 2.2 mukaisesti edes ryhmän asiakkaiden keskiarvoa käsiteltäessä ei asiakkaiden välistä hajontaa huomioida millään tavalla: keskiarvoa analysoitaessa asiakkaita oletetaan olevan paljon ja niiden oletetaan olevan melko

samanlaisia, ja toisaalta summaa käsiteltäessä ei asiakkaiden välistä hajontaa tarvitse ottaa huomioon edes periaatteessa.

Regressioanalyysiä varten tarvittava lämpötiladata luetaan tiedostosta tai tietokannasta ja käsitellään niin, että jokaiselle tunnille tallennetaan edellisen 8, 16, 24, 42 ja 40 tunnin lämpötilakeskiarvo. Tämä alustus tehdään lämpötiladataa luettaessa, joten laskentaa tehtäessä ei tarvitse enää käsitellä lämpötiladataa. Jos tunneittaista lämpötiladataa ei ole käytettävissä, voidaan laskenta tehdä saman päivän, edellisen päivän ja näiden keskiarvon perusteella. Päivänpituusdata on laskennallinen (kts. kappale 5.1), joten sähkönkulutusmittausdatan lisäksi ainut tarvittava lähtötietodata on lämpötila.

Kahden selittävänä muuttujan regressioanalyysi esitetään kappaleessa 5.2. Regressioanalyysin jälkeen saadaan kertoimet  $b_0$ ,  $b_1$  ja  $b_2$ , joiden avulla voidaan ennustaa tulevaisuuden sähkönkulutusta. Kertoimet lasketaan kunkin käyttäjän määrittelemän päiväryhmän jokaiselle 24 tunnille.  $B_0$  on vakio-osa joka kuvaa kunkin tunnin vakiona pysyvää osaa, ja  $b_1$  ja  $b_2$  ovat lämpötilan ja päivänpituuden kertoimet joiden avulla niiden vaikutus saadaan otettua huomioon. Kertoimet lasketaan kappaleessa 5.2 esitetyllä tavalla ja niiden käyttö määräytyy kappaleessa 5.4 esitettyjen periaatteiden mukaisesti: voidaan käyttää kahta selittävää tekijää, vain yhtä tai huonosti ennustettavan asiakasryhmän tapauksessa joillekin tunneille voidaan käyttää vain keskiarvoa ja keskihajontaa. Kertoimien lisäksi lasketaan myös sovituksen hyvyttä kuvaava  $R^2$  kappaleessa 5.2 esitetyllä tavalla, ja sille voidaan asettaa vaadittava minimiarvo.

Kun regressioanalyysistä on saatu ennustamisen mahdollistavat kertoimet, verrataan alkuperäisen mittausdatan arvoja niihin sähkönkulutusarvoihin jotka olisi saatu käyttämällä ennustetta ja mitattua lämpötiladataa sekä laskennallista päivänpituusdataa. Näin saadaan laskettua residuaalit, eli toteutuneen ja ennustetun kulutuksen erotukset jokaiselle tunnille. Näille residuaaleille lasketaan hajonta niin, että päiväryhmän jokaiselle 24 tunnille tulee laskettua oma keskihajonta. Residuaalien on oletettu seuraavan normaalijakaumaa, jolloin niistä lasketun keskihajonnan avulla voidaan määrittää haluttu luottamustaso kaavan 5.12 mukaisesti. Hajontaa ja normaalijakaumaoletuksen toteutumista käsitellään kappaleessa 4. Jos residuaalien hajonta ei vastaa määriteltäviä normaalijakaumankaltaisuuden vaatimuksia, voidaan siitä ilmoittaa käyttäjälle. Tällöin on oletettavaa että yllättävän suuria arvoja esiintyy enemmän kuin mitä normaalijakauma ennustaa.

Edelle esitetyn analyysin jälkeen pystytään arvioimaan minkä tahansa vuoden tunnin sähkönkulutukselle ennuste. Käyttäjän tulee valita haluttu ajankohta ja lämpötila, esimerkiksi -30 asteen arkipäivä kello 12 tammikuun alussa. Ohjelma valitsee oikean tunnin ja laskee päivänpituuden (tässä esimerkiksi päivämäärän 5.1. päivänpituus on riittävän tarkka arvio). Päivänpituus- ja lämpötila-arvo sijoitetaan kaavaan 5.12. Tämän lisäksi kaavaan sijoitetaan analyysistä saadut kello 12 kertoimet  $b_0$ ,  $b_1$  ja  $b_2$  sekä laskettu residuaalien keskihajonta  $e$ . Käyttäjän määrittelemän luottamustason mukaan valitaan  $k$ , ja näin saadaan haluttu ennuste.

Ennusteen laskemisen lisäksi käyttäjää voidaan varoittaa, jos residuaalien hajonnan analyysin aikana on ilmennyt liian paljon liian suuria arvoja: näin käyttäjä tietää että annetulla kulutusennusteella ei saavuta 95 % varmuutta. 95 % varmuus tarkoittaisi että 97,5 % varmuudella kyseisen tunnin sähkönkulutus ei ylittäisi laskettuna ennustearvoa. Mahdollisten ei-normaalijakautuneiden residuaalien lisäksi on huomioitava että tässä esitetty 95 % luottamustaso on laskettu suhteessa alkuperäiseen dataan. Käytännössä tämä tarkoittaa että asiakasryhmän oletetaan käyttäytyvän ennustettavassa tilanteessa täsmälleen samalla tavalla kuin alkuperäisen mittausdatan tilanteessa. 95 % todennäköisyys ei siis ota millään tavalla huomioon sitä, että esimerkiksi suoraa sähkölämmitystä käyttävästä asiakasryhmästä puolet olisi ostanut alkuperäisen mittausdatan mittausajankohdan jälkeen ilmalämpöpumput ja näin sähkönkulutus olisi pienentynyt.

Ohjelma on yritetty tehdä mahdollisimman joustavaksi. Esimerkiksi päivien jakaminen ryhmiin voidaan määrittää käyttäjän haluamalla tavalla, ja muistissa voi olla useita tapoja suorittaa ryhmiin jako. Tässä työssä vuosi jaetaan yleensä kuuteen päiväryhmään, mutta myös kolmen lämpötila-alueen eli yhteensä yhdeksän ryhmän jakoa käytetään. Myös lämpötilan viiveen huomioon ottamista voidaan muuttaa: esimerkiksi vähentää se vain yhteen jos laskenta vie liian kauan aikaa.

Ohjelmaan on rakennettu paljon automaattisia kuvien ja erilaisten listojen muodostamismahdollisuuksia, jotta ohjelman toimintaa voidaan seurata ja esimerkiksi normaalijakaumaoletuksen toteutumista analysoida. Kuvien muodostaminen on melko joustavaa, ja pienillä muutoksilla voidaan luoda juuri halutunlaisia kuvia esimerkiksi ennusten lopputuloksen manuaalista tarkistamista varten. Erilaisia varoituksia huonosti onnistuneesta sovituksesta voidaan myös luoda, ja esimerkiksi liian suuren yllättävän korkeiden arvojen määrän ylittyessä voidaan antaa varoitus.

## 5.1 Päivänpituuden laskeminen

Päivän pituus tunteina tiettyinä päivinä tietyssä paikassa voidaan laskea kaavalla [1]

$$D = 24 - \frac{24}{\pi} \cos^{-1} \left( \frac{\sin \frac{0,8333\pi}{180} + \sin \frac{L\pi}{180} \sin p}{\cos \frac{L\pi}{180} \cos p} \right), \quad (\text{Kaava 5.1})$$

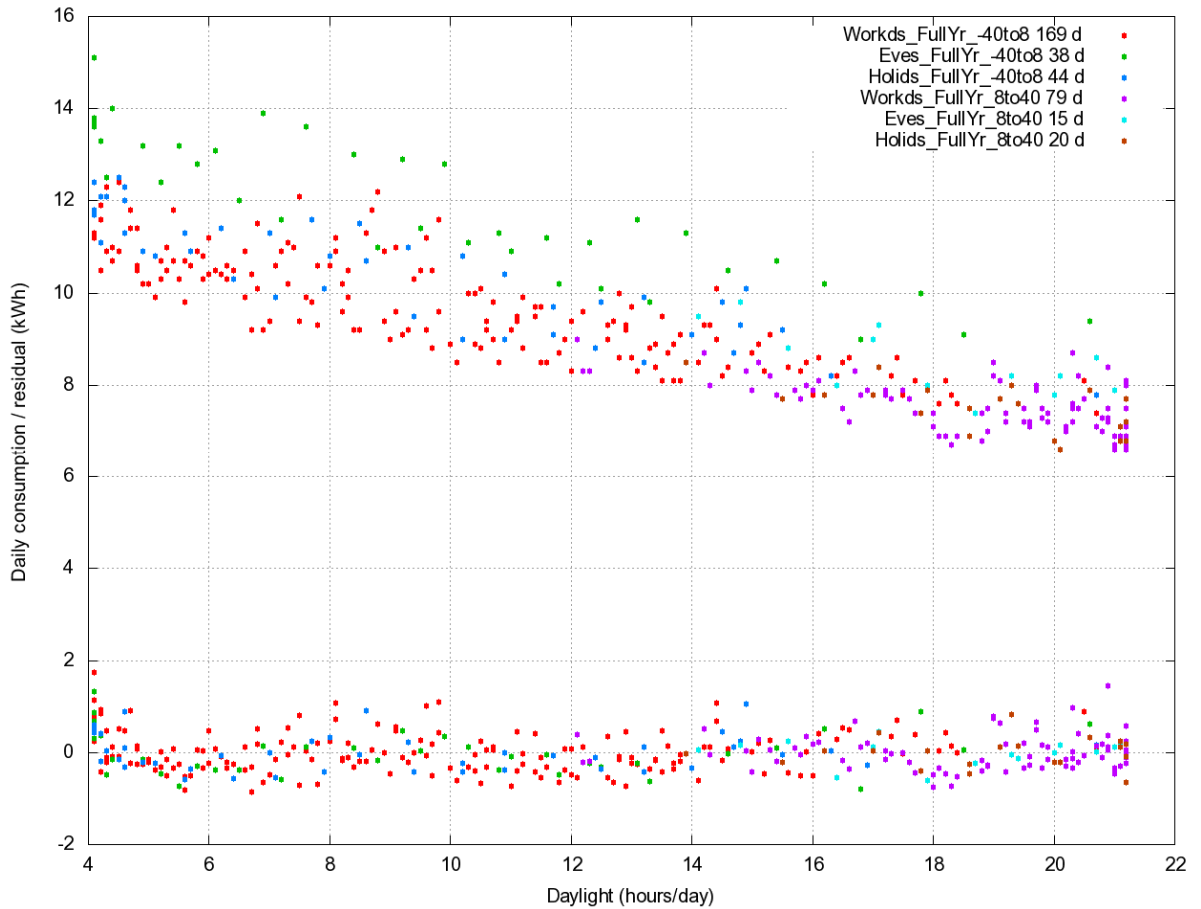
missä L on leveysaste asteissa.. Kerroin 0,8333 kertoo että päivän pituus on määritelty siitä kun auringon yläosa on silmin nähden samassa tasossa kuin horisontti. Kaavassa 5.1 p on määritelty seuraavasti [1]

$$p = \sin^{-1} \left[ 0,39795 \cos \left[ 0,2163108 + 2 \tan^{-1} \left[ 0,9671396 \tan(0,00860(J - 186)) \right] \right] \right], \quad (\text{Kaava 5.2})$$

missä J on päivän numero välillä 1..365. Kainuun leveysaste on 66,217 astetta [2]. Kaikki tässä työssä käytettä päivänpituusdata on laskettu tässä esitetyllä kaavalla, eli mitään mittaustietoa ei ole käytetty. Päivänpituus lasketaan arvona tietylle päivälle, joten tunneittaista arvoa ei saada.

Tässä työssä kaikille vuorokauden tunneille käytetään samaa päivänpituuden arvoa. Ei voida puhua sinänsä valoisuuden vaikutuksesta, koska yöllä kello yhden aikaan tiettyinä päivinä regressioanalyysiä laskettaessa käytetään samaa päivänpituuden arvoa kuin päivällä kello 17, vaan tarkoitus on löytää tietoa millä tunnilla päivänpituuden muutoksella on vaikutusta pitkällä aikavälillä: kello yhden kulutus ei juuri muutu päivän pituuden funktiona vuoden aikana, mutta kello 17 kulutus muuttuu. Matemaattisen analyysin antaman päivärakenteen loogisuutta tarkastellaan kappaleessa 6.1.

Öljy- ja kaukolämpöasiakkaiden ei oletuksena kuuluisi käyttää sähköä lämmitykseen. Kuitenkin myös öljylämmittäjille saadaan matemaattisessa tarkastelussa korrelaatio lämpötilan ja sähkönkulutuksen välille. Tämä voi olla yhtä voimakas kuin korrelaatio päivän pituuden ja kulutuksen välillä. Tämä voi johtua matemaattisesta mallista joka ei osaa erottaa lämpötilan ja päivänpituuden vaikutusta toisistaan (muuttuvat osin samaa tahtia), mutta korrelaatio voi kuvata myös todellista tilannetta. Öljylämmittäjät saattavat käyttää autonlämmittintä tai pitää esimerkiksi erillisessä rakennuksessa peruslämpöä päällä sähköpatterin avulla.



Kuva 12: Keskimääräisen kaukolämpöasiakkaan päiväkulutus (ylempi vino joukko pisteitä), kun x-akselin arvoiksi on valittu päivän pituus tunteina. Vuosi on jaettu kahteen lämpötila-alueeseen kohdasta +8 astetta, sekä kolmeen eri päivätyyppiin: esim. *Holds\_FullYear\_-40to8* sisältää ne vuoden pyhäpäivät joiden lämpötila oli välillä -40..+8 astetta. Alempi joukko pisteitä kuvaa ennusteen tekemää virhettä eli residuaalia päiväkulutusten ennustamisessa eri päivän pituuden arvoilla.

Kuvassa 12 on esitetty keskimääräisen kaukolämpöasiakkaan sähkönkulutus päivän pituuden funktiona. Vihreällä näkyvät pisteet ovat kylmien aattopäivien päiväkulutuksia, joten ne ovat hieman korkeammalla aattopäivien yleisen korkeamman kulutuksen takia. Kuvan perusteella päivänpituus näyttäisi olevan kaukolämpöasiakkaalle suhteellisen hyvä selittävä tekijä, ja yhteys on hyvin lineaarinen. Liitteessä 2 on esitetty vastaava kuva öljylämmittäjällä, mutta siinä vuosi on jaettu kolmeen lämpötila-alueeseen.

## 5.2 Kahden muuttujan regressioanalyysi

Erikoistyössäni [5] käsiteltiin ainoastaan lämpötilan vaikutusta kulutukseen. Huomattiin että sellaisillakin asiakkaila jotka eivät käytä sähköä lämmitykseen, voi olla melko voimakas matemaattinen riippuvuus lämpötilan ja sähkönkulutuksen välillä. Tämä johtuu luultavasti siitä että lämpötila ja päivänpituus muuttuvat osittain samaan tahtiin, vaikka niiden vuosittaisten huippujen välillä onkin melkein kahden kuukauden ero (nähdään kuvissa 1 ja 2). Tämän vuoksi haluttiin lisätä mahdollisuus ottaa myös päivänpituus sähkönkulutusta selittäväksi tekijäksi. Se lasketaan kaavan 5.1 avulla.

Regressioanalyysi tehdään pienimmän neliösumman (PNS) menetelmällä. Selittävien ja selitettävän välillä oletetaan lineaarinen suhde. Kahden selittävän muuttujan paras sovitussuhteessa selitettävään muuttujaan saadaan ratkaisemalla [9]

$$(X^T X)\bar{b} = X^T \bar{y}, \quad \text{missä} \quad (Kaava 5.3)$$

$$\bar{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{bmatrix} \text{ ja } \bar{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

missä y-vektori sisältää kaikki selitettävän muuttujan mitatut arvot ja X-matriisi sisältää ensimmäisen ykkössarakkeen jälkeen selittävien muuttujien mitatut arvot kukin omassa sarakkeessaan. B-vektori sisältää lopputuloksen, eli ne kertoimet joiden avulla saadaan PNS-menetelmän mukainen paras sovitus:  $b_0$  on vakio-osa ja  $b_1$  sekä  $b_2$  ovat selittävien muuttujien kertoimet.

X-matriisin ja sen transpoosin tulo on auki laskettuna

$$X^T X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{bmatrix} = \begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum (x_1 x_2) \\ \sum x_2 & \sum (x_1 x_2) & \sum x_2^2 \end{bmatrix},$$

jolloin kaava 5.3 saadaan muotoon



$$\begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum (x_1 x_2) \\ \sum x_2 & \sum (x_1 x_2) & \sum x_2^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum (y x_1) \\ \sum (y x_2) \end{bmatrix}. \quad (\text{Kaava 5.4})$$

Summaus kulkee aina ykkösestä  $n$ :ään ja  $n$  on mittauspisteiden määrä niin että yhteen mittauspisteeseen kuuluu arvot  $x_{1i}$ ,  $x_{2i}$  ja  $y_i$ . Summatermeistä nähdään, että dataa läpi käytäessä pitää koko ajan laskea kunkin selittävän ja selitettävän muuttujan summan lisäksi selittävien muuttujien neliösummaa ( $x_{1i}^2$  ja  $x_{2i}^2$ ) sekä kunkin mittauspisteen selittävien ja selitettävän muuttujan tulokombinaatioiden ( $x_{1i}x_{2i}$ ,  $x_{1i}y_i$  ja  $x_{2i}y_i$ ) summaa.

Avaamalla kaavan 5.4 sisältämät kolme yhtälöä voidaan ratkaista kertoimet kahden muuttujan parhaiten selittävää yhtälöä varten:

$$b_0 = \frac{(\sum x_1^2 \sum x_2^2 - (\sum (x_1 x_2))^2) \sum y - (\sum x_1 \sum x_2^2 - \sum x_2 \sum (x_1 x_2)) \sum (y x_1)}{n \sum x_1^2 \sum x_2^2 - n (\sum (x_1 x_2))^2 - (\sum x_1)^2 \sum x_2^2 + 2 \sum x_1 \sum x_2 \sum (x_1 x_2) - (\sum x_2)^2 \sum x_1^2} \\ - \frac{(- \sum x_1 \sum (x_1 x_2) + \sum x_2 \sum x_1^2) \sum (y x_2)}{n \sum x_1^2 \sum x_2^2 - n (\sum (x_1 x_2))^2 - (\sum x_1)^2 \sum x_2^2 + 2 \sum x_1 \sum x_2 \sum (x_1 x_2) - (\sum x_2)^2 \sum x_1^2} \quad (\text{Kaava 5.5})$$

$$b_1 = \frac{- \sum y x_1 \sum x_2^2 + \sum (x_1 x_2) \sum (y x_2) + (\sum x_1 \sum x_2^2 - \sum (x_1 x_2) \sum x_2) b_0}{- \sum x_1^2 \sum x_2^2 + (\sum (x_1 x_2))^2} \quad (\text{Kaava 5.6})$$

$$b_2 = \frac{\sum (y x_2) - \sum x_2 b_0 - \sum (x_1 x_2) b_1}{\sum x_2^2} \quad (\text{Kaava 5.7})$$

Näissä kaavoissa tarvittavat summatermit voidaan laskea melko nopeasti dataa käsiteltäessä, eikä data tarvitse käydä läpi kuin kerran.

Kahden muuttujan korrelaation voimakkuutta kuvaava Pearsonin korrelaatiokerroin voidaan laskea käyttäen hyväksi edellä esitettyjä summa-termejä seuraavasti [3]

$$r_{yx} = \frac{n \sum (xy) - \sum x \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}}, \quad (\text{Kaava 5.8})$$

missä  $r$ :n alaindeksi tarkoittaa että  $x$  selittää  $y$ :tä eli tämä kertoo yhden selittävän tekijän vaikutuksen selitettävään, mutta ei ota huomioon selittävien tekijöiden päällekkäistä vaikutusta. Näin laskettuja korrelaatiokertoimia ei voida verrata toisiinsa, vaan pitää laskea beta-kertoimet ([4] sivu 36, pdf-sivu 39), jotka voidaan kahden selittävän muuttujan tapauksessa laskea kaavalla

$$\beta_{yx.z} = \frac{r_{yx} - r_{yz} r_{xz}}{1 - r_{xz}^2}, \quad (\text{Kaava 5.9})$$

missä betan alaindeksi tarkoittaa että lasketaan  $x$ :n vaikutus  $y$ :hyn kun myös  $z$  on  $y$ :tä selittävä tekijä. Tästä kaavasta saadaan tieto siitä, millä painokertoimella  $x$  selittää  $y$ :tä.

Beta-kertoimet kertovat missä suhteessa lämpötila ja päivän pituus selittävät kulutusta. Näiden yhteisvaikutuksen selittävyden hyvyys voidaan laskea kaavasta ([4] sivu 33, pdf-sivu 36)

$$R^2 = \sum \beta r, \quad (\text{Kaava 5.10})$$

eli lasketaan kaikkien selittävien muuttujien beta-kertoimen ja korrelaatiokertoimen summa. Tämä kertoo sen, miten paljon selittävät muuttujat selittävät selitettävän varianssista. Esimerkiksi jos  $R^2 = 0,36$ , sanotaan että selittävät muuttujat selittävät 36 %. Tätä käytetään tässä työssä mittana sovituksen hyvydestä, mutta ennusteen on myös täytettävä kappaleessa 5.4 esitetyt ehdot.

Lopullinen kahden muuttujan sovitus on muotoa

$$y = b_0 + b_1 x_1 + b_2 x_2, \quad (\text{Kaava 5.11})$$

missä  $b$ -kertoimet lasketaan kaavojen 5.5, 5.6 ja 5.7 avulla.  $X$ -muuttujat ovat lämpötila sekä päivänpituus tunteina, ja  $b_0$  kuvaa vakio-osaa joka ei muutu selittävien muuttujien muuttuessa.

Selittävät muuttujat eivät pysty selittämään kaikkea sähkönkulutuksen muutosta. Jäljelle jäävä epävarmuus voidaan kuvata keskihajontana yksittäisten, käsiteltävään ryhmään kuuluvien, tuntien residuaaleista (mallin ennusteen ja alkuperäisen mittausdatan ero). Kun residuaalien oletetaan seuraavan normaalijakaumaa ja olevan riippumattomia selittävistä muuttujista, kaava 5.11 saa epävarmuustermiin sisältävänä muodon

$$y = b_0 + b_1x_1 + b_2x_2 \pm k \cdot e, \quad (\text{Kaava 5.12})$$

missä  $e$  on residuaalien keskihajonta ja  $k$ :n avulla voidaan määritellä haluttu luottamustaso, esimerkiksi 95 % luottamustasolla  $k=2$ . Hajontaa käsitellään tarkemmin kappaleessa 4.

### 5.3 Ohjelman toiminnasta

Tässä työssä esitetyt ennusteet on laskettu erikseen tätä varten tehdyllä C++ ohjelmalla. Oliopohjaista rakennetta on käytetty kaikissa ohjelman osissa. Ohjelma lukee tekstitiedostosta yksittäisten asiakkaiden vuoden tunneittaiset datat ja kokoaa näistä yhden vuoden mittaiset aikasarjan joko kaavan 2.1 tai kaavan 2.2 mukaisesti, riippuen käsitelläänkö tietyn asiakastyypin keskiarvoa vai asiakasryhmän summaa. Tämä aikasarja voidaan jakaa erilaisiin päiväryhmiin, joille kaikille tehdään oma ennuste.

Yksittäisiä asiakkaita luettaessa voidaan määritellä tiettyjä ehtoja, jotka luettavien aikasarjojen tulee täyttää. Näin voidaan jättää pois tietyt asiakkaat tai antaa varoitus mittausdatan puutteellisuudesta. Esimerkiksi voidaan vaatia että asiakkaalta tulee löytyä dataa kaikilta vuoden päiviltä tai että yli 5 % päivistä ei saa puuttua yhtään tunneittaista mittausdataa. Tällä hetkellä ohjelmaa ottaa laskentaan mukaan kaikki ryhmään määritellyt asiakkaat, mutta antaa varoituksen puuttuvista mittauksista ja kertoo kuinka monelta päivältä löytyy mittauksia.

Tällä hetkellä ohjelma lukee mittaustiedot tiedostosta, mutta tämä voidaan muuttaa helposti niin että luenta tapahtuu suoraan tietokannasta. Ohjelma tallentaa ainoastaan luettujen asiakkaiden summan, mikä pidetään käyttömuistissa. Keskiarvoa laskettaessa tämä summa vain jaetaan asiakkaiden määrällä, eli eri asiakkaiden välistä hajontaa ei oteta millään lailla huomioon. Ohjelma on suunniteltu pääasiassa asiakkaiden summan käsittelyyn, jolloin summan hajontaa laskettaessa ei tarvitse välittää yksittäisten asiakkaiden hajonnasta. Tietyn asiakasryhmän keskiarvoa käsiteltäessä lasketaan hajonta, mutta tämä tarkoittaa eri päivien välistä hajontaa.

Päivien tallentamista varten on varattu 366-paikkainen taulukko jolloin tiettyyn päivämäärään päästään käsiksi suoraan koska jokaisella päivämäärällä on tietty paikka. Päivämäärän 29.2. paikka jätetään tyhjäksi, jos kyseessä ei ole karkausvuosi. Ohjelma on tehty olio-ohjelmoinnin periaatteiden mukaan joten päivä on olio, samoin kuin tunti. Jokaiseen päivään kuuluu 24 tuntia, joten tietyn päivän tiettyyn tuntiin päästään helposti käsiksi. Tietyille tunnille tallennetaan aina

kulutuslukema, mutta tämän lisäksi tilaa saatetaan tarvita ennusteen antamaa kulutusta varten, jos se halutaan tallettaa, sekä ennusteen ja mitatun arvon välistä virhettä varten jotta voidaan tutkia virheen hajontaa.

Vaikka ohjelmassa tällä hetkellä tallennetaan vain yksi vuoden mittainen aikasarja kulutusdataa, on ohjelman rakenteessa varauduttu usean aikasarjan pitämiseen muistissa. Näin voidaan tutkia esimerkiksi usean yksittäin käsitellyn ryhmän välistä korrelaatiota. Lämpötiladata tallennetaan erikseen, mutta se vie enemmän tilaa. Lämpötilasta tallennetaan jokaisen tunnin arvo, mutta tämän lisäksi lämpötiladata käsitellään niin että jokaiselle tunti tietää myös menneiden tuntien keskiarvon. Näitä keskiarvoja voidaan tallentaa erilaisia määriä, mutta tällä hetkellä niitä tallennetaan viisi kappaletta: 8, 16, 24, 32 ja 40 edellisen tunnin keskiarvo. Näiden tietojen avulla voidaan nopeasti laskea viisi erilaista lämpötilakorrelaatiota jotka yrittävät ottaa huomioon lämpötilariippuvuuden hitauden.

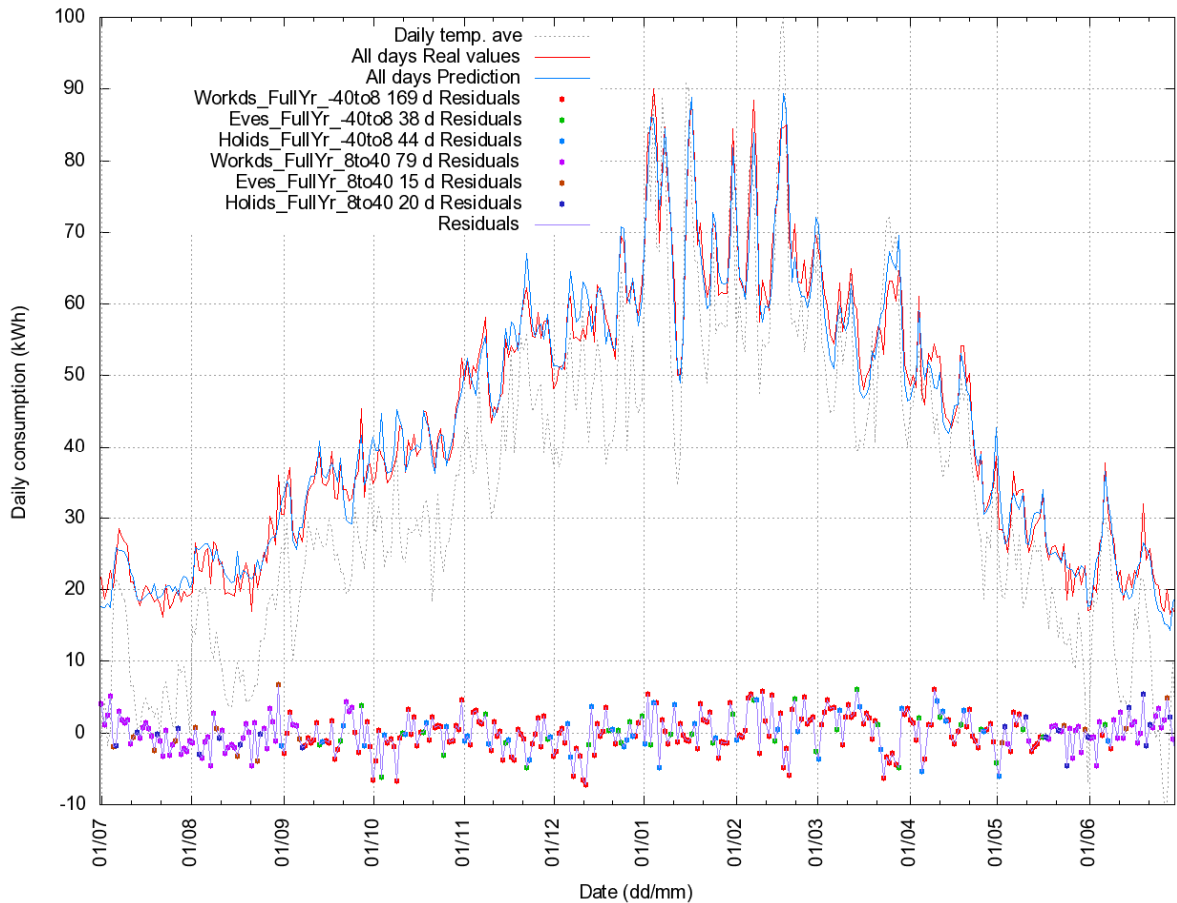
Vuoden päivät pitää pystyä käsittelemään ryhmissä. On kuitenkin turha kopioida koko päivän muistiin kirjoitettua dataa, joten tämä on ratkaistu käyttämällä filteri-luokkaa joka sisältää linkit sen määrittelyt täyttyviin päiviin. Filtteriin voidaan määritellä päivä tyyppi, päivän lämpötilan keskiarvon ala- ja yläraja sekä mahdollisesti muita määritteitä. Päivän tyyppi voi olla esimerkiksi vain arkipäivät, vain pyhäpäivät tai vain tiistait. Päivän tyyppi selvitetään päivämäärän avulla käyttämällä viikonpäivän laskemiseen kaavaa [16]

$$h = \left( q + \frac{26(m+1)}{10} + K + \frac{K}{4} + \frac{J}{4} + 5J \right) \bmod 7, \quad (\text{Kaava 5.13})$$

missä  $q$  on kuukauden päivän numero,  $m$  on kuukausi (3 = maaliskuu, 4 = huhtikuu, ..., 14 = helmikuu),  $K = y \bmod 100$  ja  $K = y/100$ , missä  $y$  on vuoden numero. Kuukauden numeroinnissa tammi- ja helmikuu kuuluvat edelliseen vuoteen, eli silloin vuoden numerosta pitää vähentää yksi. Kaavasta saatu viikonpäivä on numeroitu niin, että 0 = lauantai, 2 = sunnuntai, ..., 6 = perjantai.

Kaavan 5.13 avulla saadaan vain osittain erotettua arki-, aatto- ja pyhäpäivät. Lisäksi tarvitaan tieto vuosittaisista aatto- ja pyhäpäivistä jotka eivät osu lautoille tai sunnuntaille. Tämä tieto on haettu wikipediasta [17]. Osa pyhäpäivistä määräytyy pääsiäisen perusteella ja pääsiäisen päivämäärä taas voi vaihdella [18], joten tämä otetaan huomioon.

Ohjelmassa voi jakaa vuoden joustavasti erilaisiin päiväryhmiin ja näitä ryhmiä voi luoda lisää. Tämä pitää kuitenkin tehdä niin, että kaikki päivät kuuluvat johonkin ryhmään mutta vain yhteen ryhmään. Tässä työssä on käytetty jakoa kolmeen päivätyyppiin (arki-, aatto- ja pyhäpäivät) sekä kahteen lämpötila-alueeseen (päivän lämpötilakeskiarvo yli tai alle +8 astetta).

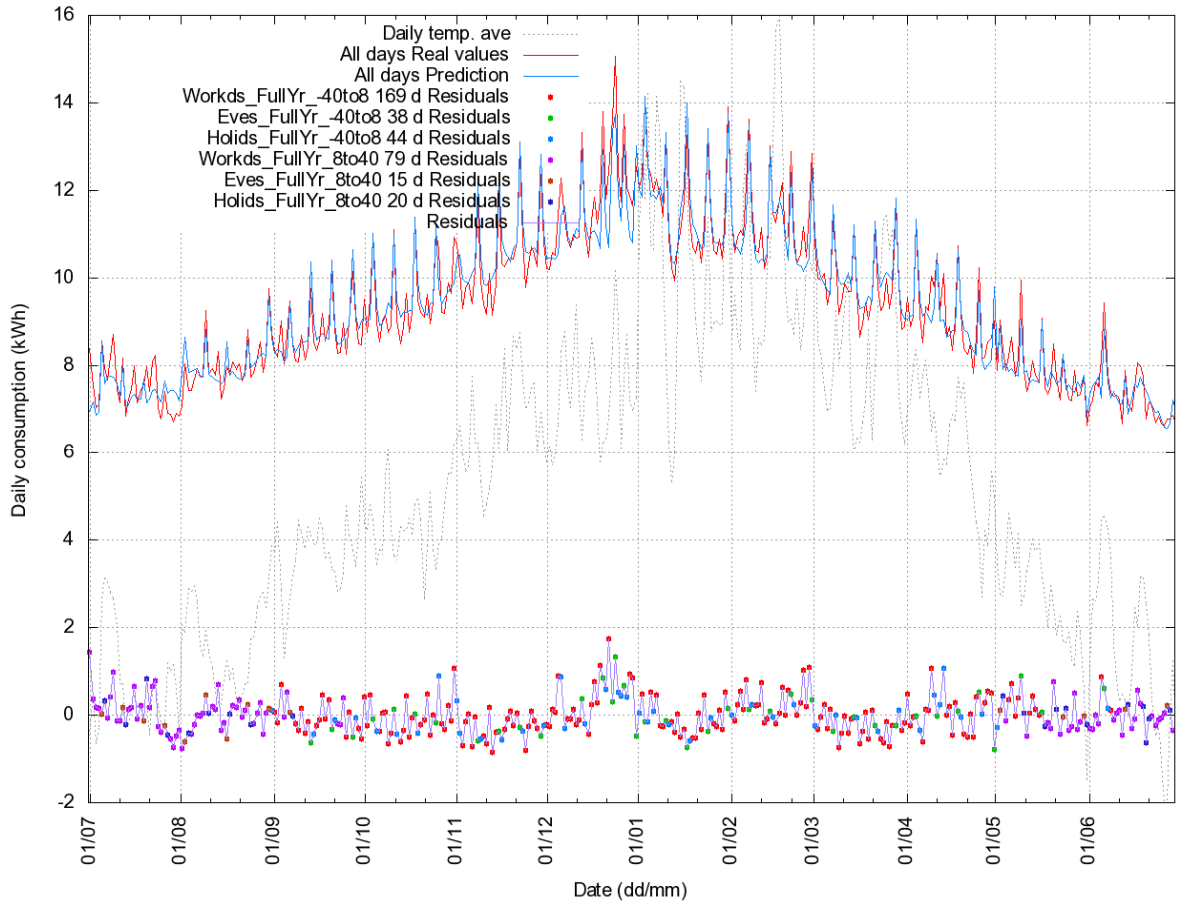


*Kuva 13: Keskimääräisen varaavan sähkölämmittäjän mitattu (punainen) ja ennustettu (sininen) päiväkulutus vuoden aikana, kun vuosi on jaettu kuuteen päiväryhmään. Kuvan alaosassa on kuvattuna ennusteen tekemän virhe eli residuaali vuoden eri päivinä.*

Kullekin päiväryhmälle lasketaan kahden muuttujan regressioanalyysi. Tämä lasketaan niin moneen kertaan, että regressio tulee laskettua kaikilla määritellyillä tavoilla. Valoisuus lasketaan aina samalla tavalla, mutta lämpötiladataa voidaan käyttää eri tavoilla: tässä ohjelmaversiossa regressio lasketaan 8, 16, 24, 32 ja 40 edellisen tunnin keskiarvoilla. Tämän takia regressio pitää laskea viiteen kertaan.

Regressiotuloksista tallennetaan vain parhaan tuloksen antanut, eli kaavan 5.10 mukaisesti suurimman arvon antanut regressiosovitus. Päivänpituuden ja lämpötilan vaikutuksille voidaan

kuitenkin asettaa rajat esimerkiksi kappaleessa 5.4 esitetyllä tavalla. Regressioanalyysin vaatimiin laskutoimituksiin tarvittavat summatermit saadaan käymällä data läpi yhden kerran. Residuaaleja eli virheitä laskettaessa pitää kuitenkin data käydä vielä toisen kerran läpi.



*Kuva 14: Keskimääräisen kaukolämpöasiakkaan mitattu (punainen) ja ennustettu (sininen) päiväkulutus vuoden aikana, kun vuosi on jaettu kuuteen päiväryhmään. Kuvan alaosassa näemme ennusteen tekemän virheen eli residuaalin vuoden eri päivinä.*

Kun regressio on tehty ja on saatu parhaimman tuloksen antaneet regressiokertoimet, ne tallennetaan muistiin ja niitä käytetään ennustettaessa tulevaisuuden kulutusta. Yhdeltä käsitellyltä ryhmältä tulee tallettaa tässä ohjelmaversiossa  $6 \cdot 24 \cdot (3+1) = 576$  arvoa, jotta voidaan tehdä arvio mille tahansa tunnille: 6 tulee päiväryhmien määrästä, 24 tuntien määrästä, 3 tarvittavista kertoimien määrästä eli  $b_0$ ,  $b_1$  ja  $b_2$ , ja +1 siitä että lämpötilaa käytettäessä pitää tietää mikä testatusta 5 lämpötilakeskiarvosta antoi parhaan tuloksen. Tämän 576 arvon lisäksi saatetaan kuitenkin haluta tallentaa vielä virheen hajonta, jolloin arvoja tulisi tallentaa yhteensä  $576 + 24 \cdot 6 = 720$  kappaletta. Jos tämä on liikaa, voidaan määrää vähentää esimerkiksi laskemalla jokaisen

tunnin sijasta esimerkiksi 3 tunnin keskiarvoja tai käyttämällä 6 päiväryhmän sijasta vain muutamaa ryhmää.

Kuvissa 13 ja 14 nähdään varaavan keskimääräisen varaavan sähkölämmittäjän ja kaukolämpöasiakkaan alkuperäinen mittausdata päiväkulutuksina ajalta 1.12–28.2.2009, sekä saman datan avulla tehty ennuste. Varaavan sähkölämmittäjän kulutus muuttuu lämpötilan mukaan, eikä esimerkiksi joulun ajan korkeampi kulutus näy residuaaleissa. Kaukolämpöasiakkaan yleisesti pienemmässä kulutuksessa joulu näkyy hieman suurempana virheenä. Tämä oli oletettavaa ja pystytään ottamaan kunnolla huomioon vain käsittelemällä joulu täysin erikseen. Tässä 23.12 ja 24.12 on laitettu aattopäivien ryhmään. Suoran sähkölämmittäjän ja öljylämmittäjän vastaavat kuvat on esitetty liitteissä 3 ja 4. Esimerkki kolmeen lämpötila-alueeseen jaosta ja näin tehdyn ennusteen päiväkulutuksista verrattuna mitattuihin arvoihin on esitetty liitteessä 5.

## 5.4 Regressiotulosten automaattinen tarkistus

Regressiolaskentaa tehdessä tarkistetaan matemaattisen käsittelyn antamat tulokset ja muutetaan tarvittaessa laskentaa niin, että saadaan sekä matemaattisesti että fysikaalisesti järkevät tulokset. Tarkistukset tehdään painottaen luotettavuutta ja robustisuutta tarvittaessa ennusteen tarkkuuden kustannuksella. On pyritty välttämään tilannetta, jossa malli saattaisi antaa epärealistisia tuloksia.

Oletuksena regressio lasketaan kahden selittävän muuttujan, päivän pituuden ja lämpötilan, avulla. Tämän lisäksi lasketaan myös erikseen kummankin selittävän muuttujan regressio ilman toista muuttujaa tarkistusta varten. Jos toinen muuttujista ei täytä määriteltäviä vaatimuksia, se jätetään pois ja regressio määritellään uudestaan vain yhtä selittävää muuttujaa käyttäen. Jos kumpikaan muuttujaa ei täyty vaatimuksia, käytetään kyseisin tunnin ennusteena keskiarvoa johon lämpötila tai päivän pituus ei vaikuta mitenkään. Kummankin selittävän tekijän selityksen hyvyys lasketaan kaavalla 5.8, korottamalla  $r$  potenssiin kaksi. Sen lisäksi lasketaan näiden yhteisvaikutus, eli koko sovituksen hyvyys, kaavalla 5.10.

Päivänpituuden käytölle regressiolaskennassa on asetettu kolme ehtoa. Ensimmäiset ehdot on määriteltä niiden tulosten pohjalta, mitä saadaan laskettaessa pelkän päivänpituuden vaikutusta sähkönkulutukseen. Ensinnä päivänpituuden vaikutuksen on oltava negatiivinen, eli päivän lyhentyessä kulutuksen pitää kasvaa. Toiseksi pelkän päivän pituuden avulla tehdyn analyysin

pitää pystyä selittämään kulutusta niin, että korrelaatiokertoimen itseisarvoksi saadaan 0.2 tai suurempi. Tällöin selityksen hyvyys ( $r^2$ ) on vain 0.04 eli 4 %. Vaadittua selityksen hyvyyttä voidaan säätää ohjelmassa helposti korkeammalle tasolle.

Lämpötilalle on asetettu muuten samat vaatimukset kuin päivän pituudelle, mutta lämpötila voi vaikuttaa myös vastakkaiseen suuntaan jotta kesän mahdollinen ilmastointi saadaan otettua huomioon. Sekä päivän pituuden että lämpötilan pitää myös toimia samalla tavalla kun regressio lasketaan erikseen yhdellä muuttujalla ja kun muuttujat vaikuttavat samaan aikaan. Tämä tarkoittaa että kaavan 5.11 kertoimien  $b_1$  ja  $b_2$  etumerkkien on oltava samat kuin laskettaessa yhden muuttujan regressiot erikseen. Jos lämpötila siis vaikuttaa tietyllä tavalla kun regressio lasketaan vain sen avulla, ei tämä vaikutus saa olla päinvastainen kahden muuttujan regressiossa ja sama päivän pituuden tapauksessa. Tällainen on melko harvinaista mutta epäloogista, joten se on haluttu sulkea lopputulosmahdollisuuksien ulkopuolelle.

Tässä esitettyjä vaatimuksia regressionanalyysin lopputulokselle on määritelty lähinnä ongelmien ilmentyessä. Vaaditun sovituksen hyvyyden suuruuksia voidaan muuttaa helposti. Tämän lisäksi olisi kuitenkin hyvä testata regressiosovituksien loogisuutta vielä lisää tämän työn jälkeen ja määritellä tarpeen mukaan lisää ehtoja. Näin voidaan välttyä epäloogisten matemaattisten tulosten aiheuttamilta ongelmilta.



## 6 Ennusteen analysointi

Tässä kappaleessa testataan kappaleen 5 mukaisesti tehdyn ennusteen loogisuutta ja hyvyttä eri lämmitystyyppien käyttävien asiakkaiden sähkönkulutuksen ennustamisessa. Tämän lisäksi testataan satunnaisesti valittujen pienten asiakasryhmien summan ennustamista. Näiden on tarkoitus kuvata muuntajan kuorman summaa.

Ennusteen loogisuutta tutkitaan analysoimalla eri asiakastyypin päivärakennetta. Näissä pitäisi olla isoja eroja eri lämmitystyyppien välillä. Saatua ennustetuloksia verrataan lähinnä oletukseen tietyistä asiakastyypeistä. Ennusteen hyvyttä tutkitaan keskittyen suurimpiin tuntikulutuksiin, koska niiden ennusteen tarkkuus on kaikkein tärkeintä.

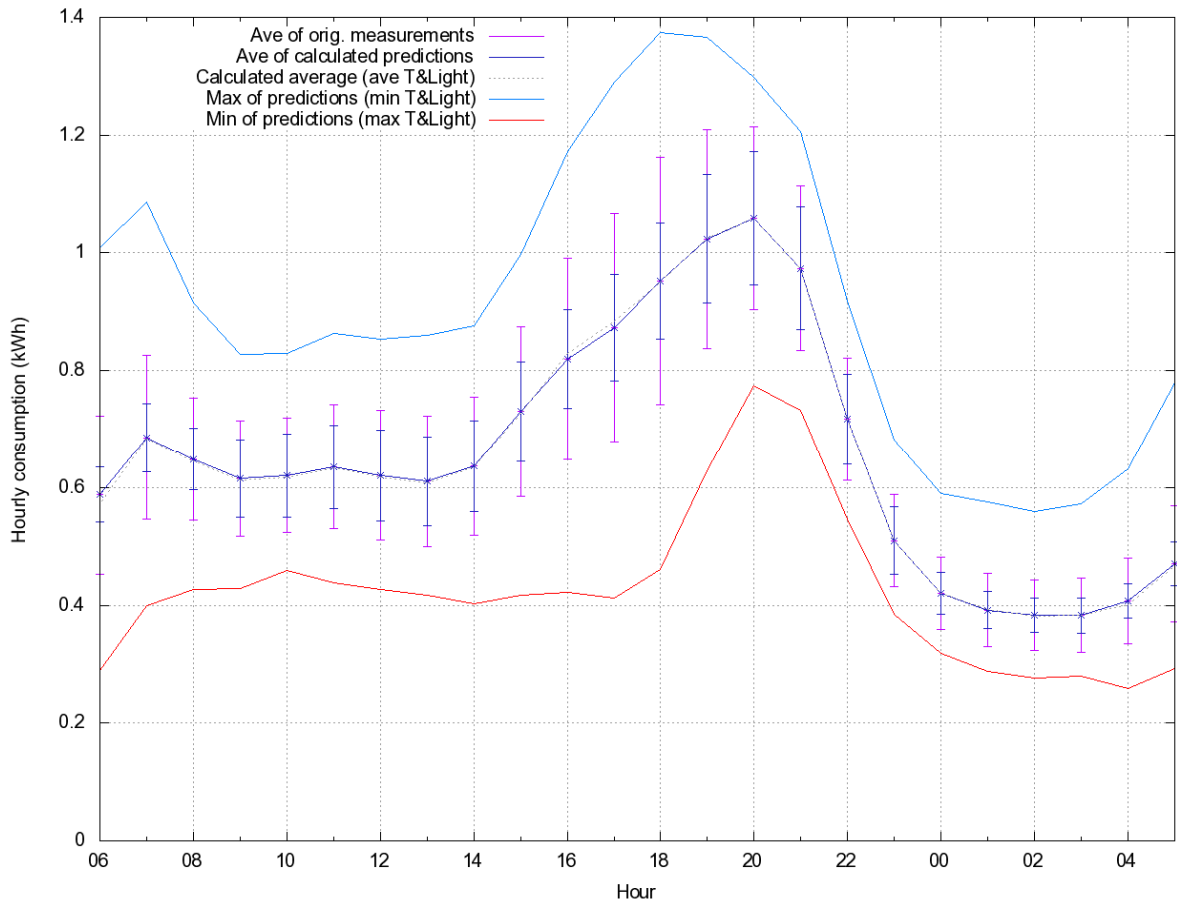
### 6.1 Eri lämmitystyyppien päivärakenteet

Jokaiselle tunnille lasketaan erikseen kahden selittävän muuttujan regressioanalyysi kaavojen 5.5, 5.6 ja 5.7 mukaisesti. Näin päivälle tulee tietty rakenne, jossa jokaisella tunnilla on sekä vakio-osuus että kertoimet joiden suuruus kertoo lämpötilan ja päivänpituuden muutoksien vaikutuksen kyseisen tunnin kulutukseen. Tässä kappaleessa tutkitaan näiden ennusteen antamien rakenteiden järkevyyttä.

Päivänpituus voi ennusteen tuloksissa vaikuttaa sähkönkulutukseen myös illalla ja yöllä. Valoisuus sinänsä vaikuttaa tietenkin vain päivällä, mutta päivänpituuden vaikutukseen luetaan kaikki kulutuksen muutos mikä muuttuu samaan aikaan päivänpituuden kanssa vuoden aikana. Matemaattinen analyysi saattaa löytää yhteyden päivänpituuden muutoksen ja sähkönkulutuksen muutoksen välille, koska lämpötilan muutos seuraa päivänpituuden muutosta kuten kuvissa 1 ja 2 nähdään. Illan kulutus voi kuitenkin myös todellisuudessa seurata päivänpituuden muutosta, koska ihmiset saattavat esimerkiksi katsoa enemmän televisiota kun ulkona on pimeää.

Kuvassa 15 nähdään öljylämmittäjän kylmien päivien päivärakenne. Tässä päivän rakenne näyttää melko loogiselta: aamulla ja illalla korkeampi kulutus ja yöllä matalampi. Kaikkien päivän tuntien alkuuperäisestä hajonnasta saatiin lämpötilan ja päivänpituuden avulla selitettyä osaa, mutta kello 16–18 hajonnasta jonkin verran enemmän kuin esimerkiksi kello 21 hajonnasta. Tämän seurauksena matalan kulutuksen aikaan (punainen käyrä) päivän huippukulutus sattuu eri tunnille

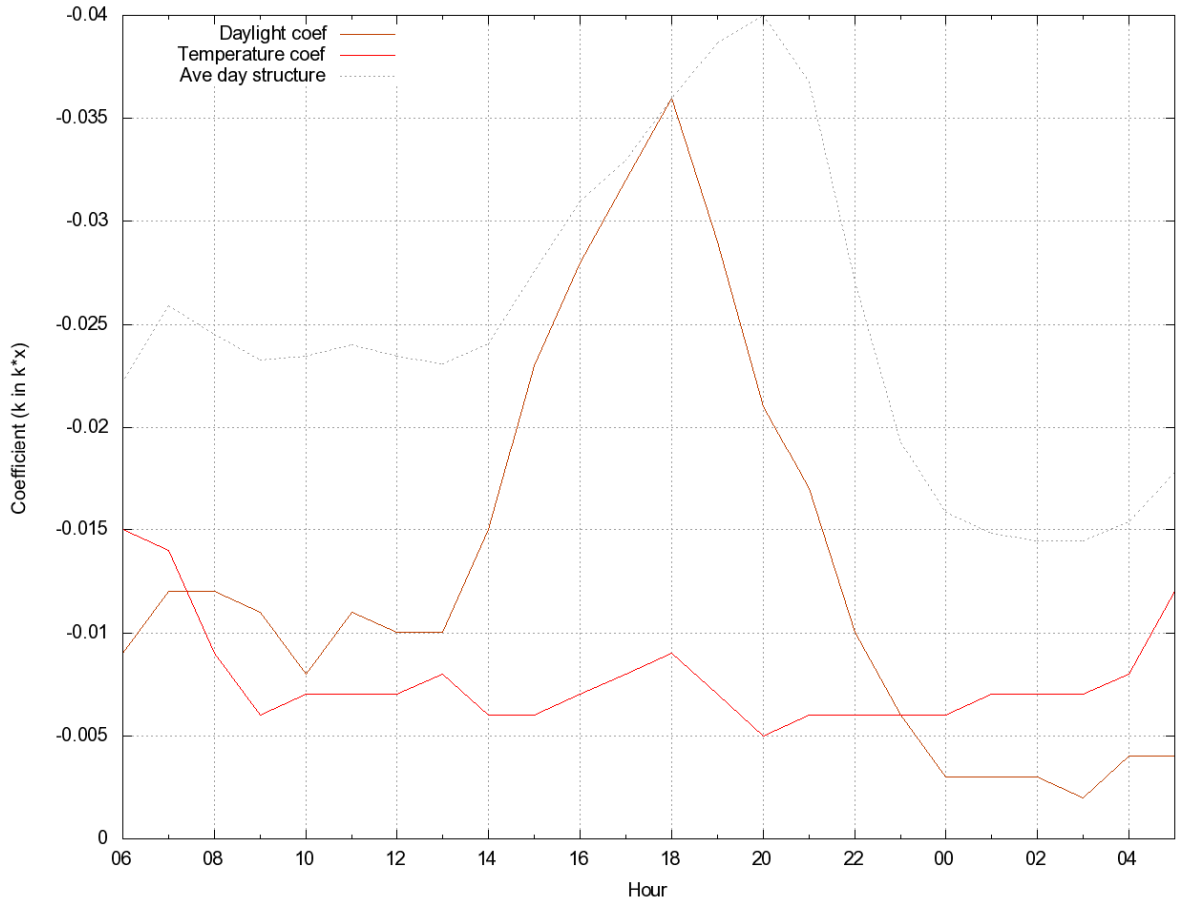
kuin korkean kulutuksen aikaan (vaaleansininen käyrä): ennustettu kulutushuippu siirtyy tunnilta 20 tunnille 18. Kello 20:n kohdalla selittämättä jäävä hajonta on kuitenkin suurempi kuin kello 18, joten se pitää ottaa huomioon luottamustasoja määritettäessä: kello 20 kulutus on siis vähemmän ennalta-arvattavaa.



*Kuva 15: Tumman sinisellä keskimääräisen öljylämmittäjän päivärakenne arkipäivänä, kun päivän keskiarvolämpötila on alle +8 astetta. Kuvaan on merkitty myös hajonta eri tunneilla ennen sovitusta tummansinisellä ja lilalla sovituksen jälkeen: pieni hajonta kuvaa hyvin onnistunutta sovitusta. Vaaleansinisellä on kuvattu mallin antama ennuste jos samaan aikaan olisi sekä erittäin kylmää että pimeää (käsitellyn aikavälin toteutuneet minimi), eli se kuvaa mallin antamaa suurinta mahdollista päiväkulutusta. Punainen käyrä kuvaa samalla periaatteella mallin antamaa minimipäiväkulutusta.*

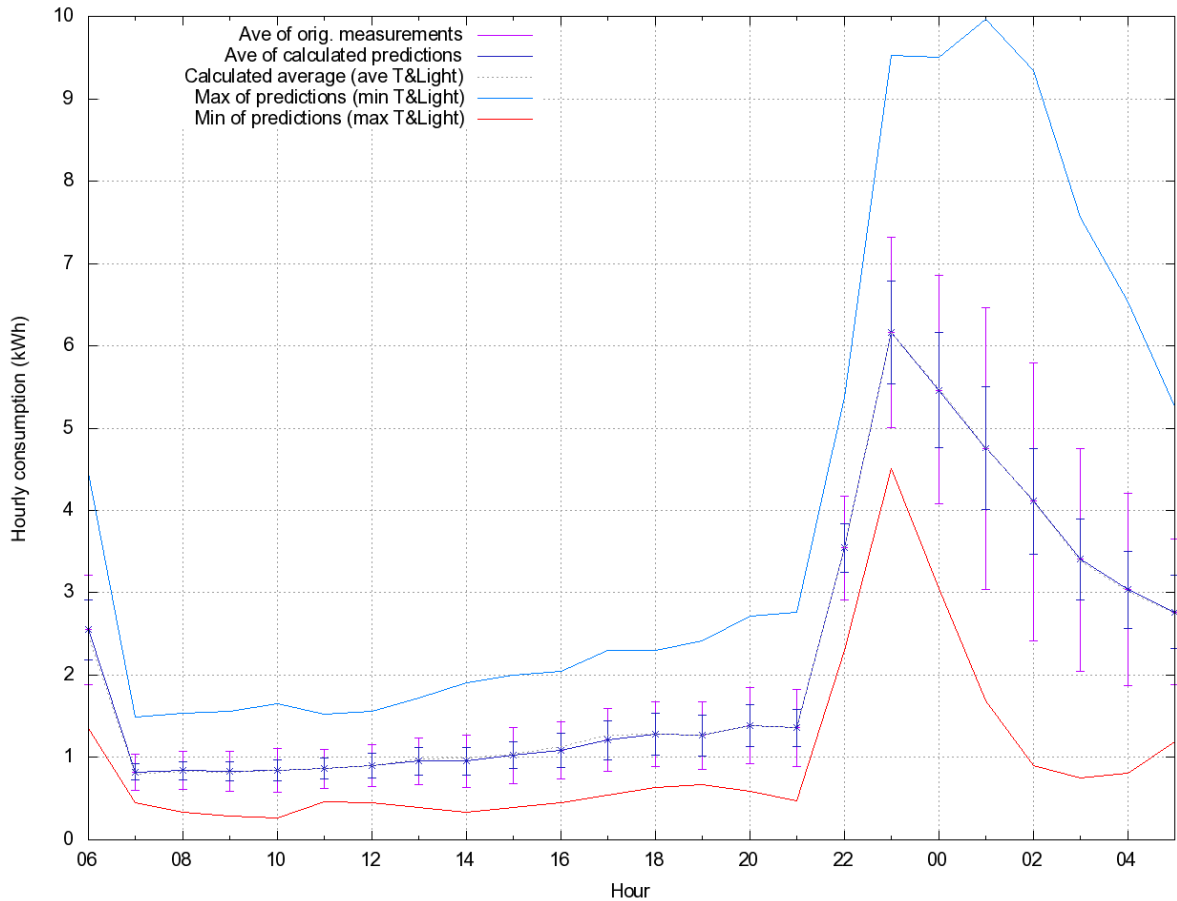
Kuvassa 15 vaaleansininen käyrä on analyysin antama ennuste jos sekä lämpötila että päivänpituus ovat minimissään, eli arvio on kulutusmaksimi. Talvella tämä tarkoittaisi että samaan aikaan on yhtä pimeä kuin 26.12. ja yhtä kylmää kuin koko talven kylmin hetki (riippuu käytetystä regressiomallista onko tämä päiväkeskiarvo vai esimerkiksi 8 tunnin keskiarvo) oli, joten käytännössä tällainen tilanne on harvinainen. Lämpötilaminimi otetaan kuitenkin toteutuneesta lämpötiladatasta, joten kylmempänä vuonna voidaan saada suurempi

tuntikulutusmaksimi. Punainen käyrä kertoo samalla periaatteella minimikulutuksen. Näiden tarkoitus on lähinnä näyttää mitkä tunnit muuttuvat eniten lämpötilan ja päivänpituuden muuttuessa.



Kuva 16: Keskimääräisen öljylämmittäjän regressionsovituksen antamat kertoimet arkipäivänä, kun päivän keskiarvolämpötila on alle +8 astetta. Oranssi ja punainen käyrä kuvaavat kertoimia  $b_1$  ja  $b_2$  kaavassa 5.11 eri kellonaikoina. Oranssi kuvaa valoisuuden vaikutusta ja punainen lämpötilan vaikutusta. Esim. klo. 18 valoisuuden selittävä vaikutus sähkönkulutukseen on suurimmillaan. Tässä esitettyjen kertoimien mukainen päivä rakenne nähdään kuvassa 15.

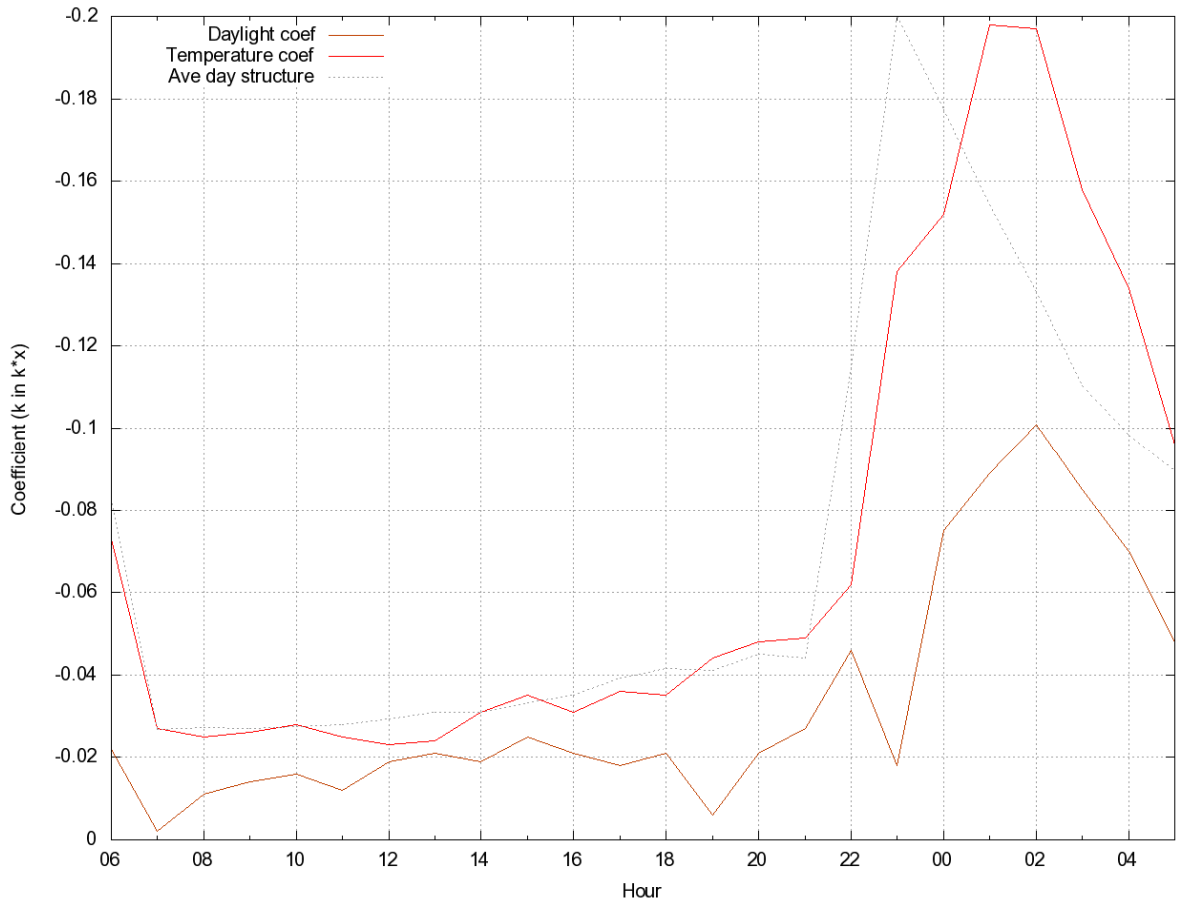
Kuvassa 16 näkyy se, selittääkö tietyn tunnin muutosta paremmin lämpötila vai päivänpituus. Kuvassa näkyvät arvot ovat kaavan 5.11 kertoimet  $b_1$  ja  $b_2$ . Kuvassa voi verrata esimerkiksi kello kolmen ja kello 18 päivänpituuden vaikutusta (oranssi käyrä), mutta ei suoraan saman tunnin päivänpituuden ja lämpötilan vaikutusta koska arvot eivät ole vertailukelpoisia keskenään. Tämä johtuu siitä, että päivänpituus ei vaihtele yhtä paljon kuin lämpötila, eli kaavassa 5.11 kertoimilla  $b_1$  ja  $b_2$  ei ole yhtä suurta vaikutusta lopulliseen kulutusennusteeseen. Kaavassa päivän pituuden vaikutus on kertoimen ja mittausarvon tulo joten pelkkä kertoimen  $b$  suuri arvo ei tee tulosta suurta.



Kuva 17: Tummansinisellä keskimääräisen varaavan sähkölämmittäjän päivärakenne arkipäivänä, kun päivän keskiarvolämpötila on alle +8 astetta. Kuvaan on merkitty myös hajonta eri tunneilla ennen sovitusta tummansinisellä ja lilalla sovituksen jälkeen: pieni hajonta kuvaa hyvin onnistunutta sovitusta. Vaaleansinisellä on kuvattu mallin antama ennuste jos samaan aikaan olisi sekä erittäin kylmää että pimeää (käsitellyn aikavälin toteutuneet minimi), eli se kuvaa mallin antamaa suurinta mahdollista päiväkulutusta. Punainen käyrä kuvaa samalla periaatteella mallin antamaa minimipäiväkulutusta.

Kuva 16 näyttää suhteellisen järkevältä: lämpötilan vaikuttaa hieman koko ajan, mutta päivän pituuden muutos ei juuri vaikuta yöllä vaan sen vaikutuksen maksimiarvo on kello 18. Kuvasta nähdään sama mikä kuvasta 15, eli kello 20 korkean kulutuksen hetkeä ei pystytty selittämään kovin hyvin vaan kulutus on jatkuvasti melko korkea ja kulutuksen muutos täytyy lukea hajonnan piikkiin. Kello 18 päivänpituuden vaikutus on arkipäivänä suurimmillaan, mutta kello 20 kulutus on keskimäärin suurimmillaan. Kello 20 kulutusta ei kuitenkaan selitä päivänpituus vaan silloin kulutus on suurta valon määrästä huolimatta. Kuvassa 15 maksimikulutuksen aikana (sininen käyrä), eli kun on pimeää, kulutus alkaa nousta jo n. kello 14–15 oletettavasti koska tällöin jo tarvitaan valoa mutta minimikulutuksen aikana (punainen käyrä) kulutus on korkeaa vasta kello 19–20 jolloin kulutus on jatkuvasti melko suurta.

Varaavan sähkölämmittäjän päivärakenne nähdään kuvassa 17. Kuvassa nähdään piikki kello 23. Tämä on yösähkön alkamisaika, ja piikki esiintyy myös suoraa sähkölämmitystä käyttävien asiakkaiden päivärakenteessa: asiakkaat voivat käyttää yösähköä esimerkiksi veden lämmittämiseen. Varaavan sähkölämmittäjän tapauksessa nähdään lämmityksen suhde energiaan eikä vain huipputehoon. Kylmimmän ja pimeimmän päivän aikana (sininen käyrä) kulutus pysyy suurin piirtein samalla tasolla tuntien 22-2 aikana lisäten näin lämmitysenergiaa vaikka huipputeho ei kasva. Tilanne näyttää tässä kuvassa melko järkevältä mutta kappaleessa 7.3 käsitellään tarkemmin varaavan sähkölämmittäjän ennustetta ja varsinkin huipputehon ennustamista sekä siihen liittyviä ongelmia. Huipputehon muutos on melko suurta kello 23. Päivänpituuden ei pitäisi vaikuttaa yöllä, mutta matemaattisissa analyysissä on joskus vaikea erottaa sen ja lämpötilan vaikutusta koska ne muuttuvat samaan aikaan vuoden kuluessa.



Kuva 18: Keskimääräisen varaavan sähkölämmittäjän regressionsovituksen antamat kertoimet arkipäivänä, kun päivän keskiarvolämpötila on alle +8 astetta. Oranssi ja punainen käyrä kuvaavat kertoimia  $b_1$  ja  $b_2$  kaavassa 5.11 eri kellonaikoina. Oranssi kuvaa valoisuuden vaikutusta ja punainen lämpötilan vaikutusta. Esim. klo. 1 yöllä lämpötilan selittävä vaikutus sähkönkulutukseen on suurimmillaan. Tässä esitettyjen kertoimien mukainen päivärakenne nähdään kuvassa 17.

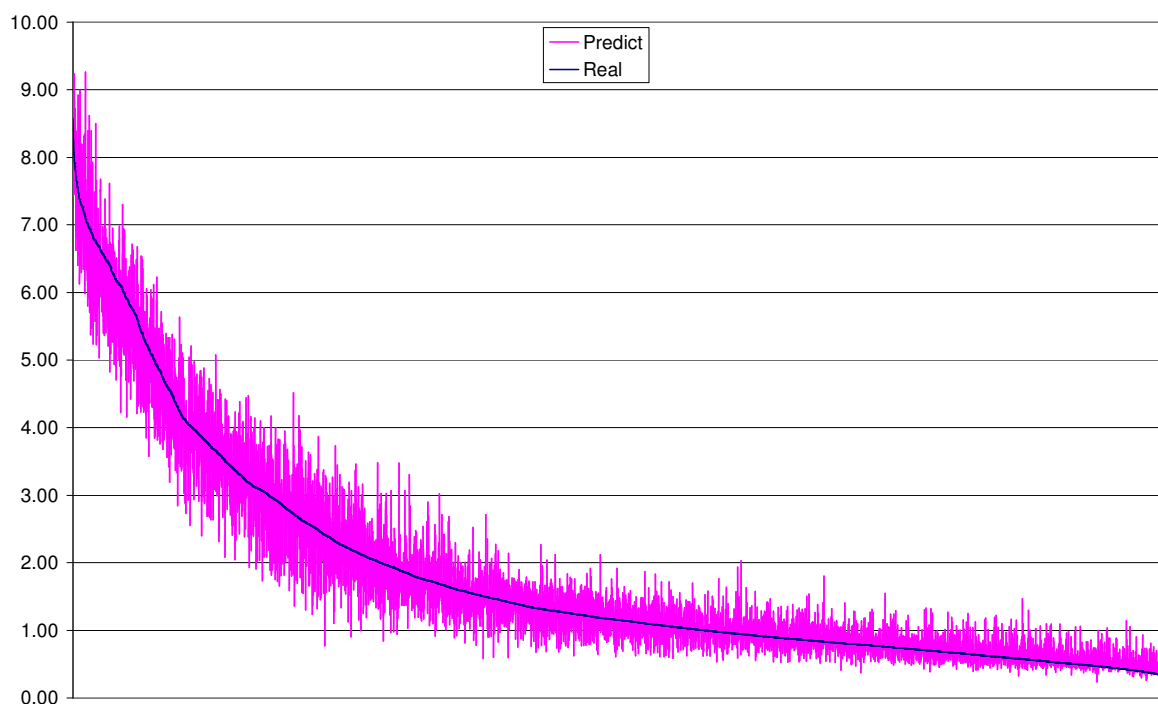
Kuvassa 18 nähdään kuvaa 17 vastaavan päiväryhmän lämpötilan ja päivän pituuden selittävää vaikutusta kuvaavat kertoimet  $b_1$  ja  $b_2$ . Huomataan että kello 23 kulutuspiikkiä ei pystytty selittämään selittävien muuttujien avulla, vaan silloin kulutus on lähes aina korkea. Aamuyön kulutus taas pystytään selittämään suurimmaksi osaksi lämpötilan avulla. Päivän pituuden suhteellisen korkea kerroin aamuyöllä on hieman yllättävä, mutta ei kohtuuton. On myös huomattava että korkea päivänpituuden kerroin  $b$  kaavassa 5.11 ei vaikuta yhtä paljon kuin lämpötilan kerroin koska päivänpituuden arvo ei muutu yhtä paljon kuin lämpötila.

Tässä käsitellyt päivä rakenteet kuvaavat eri lämmitystyyppien eroja, ja ne tuntuvat melko järkeviltä. Ne ovat kuitenkin kylmien arkipäivien pohjalta tehtyjä ja sisältävät kaikista ryhmistä eniten päiviä. Aatto- ja pyhäpäivät eivät aina anna yhtä hyvää tulosta, ja kesällä laskennallinen päivä rakenne voi olla melko sekava koska selittävät muuttujat eivät pysty selittämään sähkönkulutuksen muutosta. Esimerkit aatto- ja pyhäpäivien rakenteista nähdään liitteissä 8 ja 9 varaavalle sähkölämmittäjälle ja öljylämmittäjälle. Keskimääräisen suoran sähkölämmittäjän ja kaukolämpöasiakkaan päivä rakenteet on esitetty liitteissä 6 ja 7.

Yksi mitta päivä rakenteen hyvyydelle ja loogisuudelle on valoisuuden ja lämpötilan vaikutuksen jatkuvuus. Esimerkiksi kuvassa 18 kertoimien arvot eivät heittele ylös alas, vaan ne kulkevat melko tasaisella tahdilla ensin ylös kello 20-0 ja sitten alas aamuyöllä. Esimerkki huonosti onnistuneesta päivä rakenteen analyysistä nähdään kuvassa 22.

## 6.2 Eri lämmitystyyppien huippukulutuksen ennustaminen

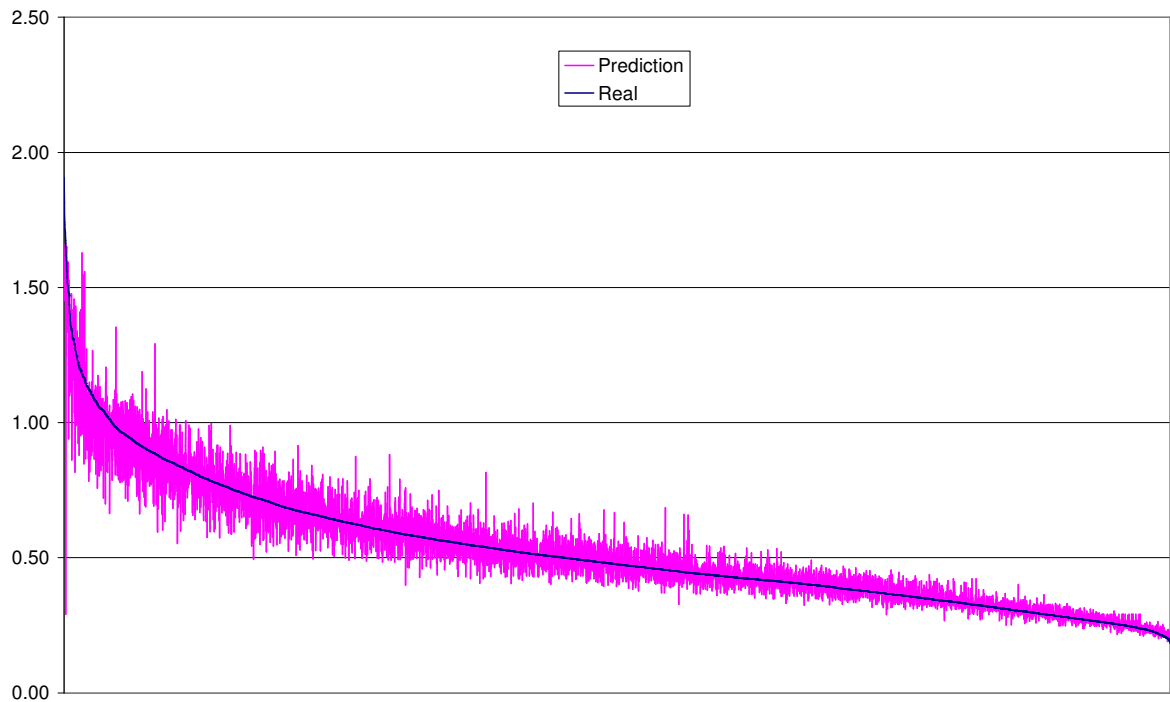
Sähkönkulutusennusteen hyvyttä arvioitaessa tärkeintä on huippukulutuksen arvion tarkkuus ja luotettavuus, koska sen mukaan verkko täytyy mitoittaa. Tämän lisäksi halutaan yleensä tietää vuosien energia sekä mahdollisesti jotain tehon pysyvyydestä, esimerkiksi huipunkäyttöaika. Tässä työssä esitellyn ohjelman avulla voidaan laskea nämä kaikki, mutta tarvittaessa myös enemmän tunnuslukuja. Kuvassa 19 nähdään keskimääräisen varaavan sähkölämmittäjän tunneittainen pysyvyyskäyrä alkuperäisen mitatun datan mukaisesti. Nähdään että kulutus on muuttunut melko hyvin ennusteen mukaisesti. Korkean kulutuksen tunnit, esimerkiksi kello 23 ja 0 kuten näemme taulukossa 3, ovat olleet ennustettavissa. Suurimman virheen tunnit ovat esimerkiksi kello 20 tunteja jolloin ennuste on kaikkein epävarmin.



Kuva 19: Keskimääräisen varaavan sähkölämmittäjän tuntikulutukset järjestettynä alkuperäisen mittauksen suuruuden mukaan sinisellä, sekä lilalla ennusteen antamat arvot samoille tunneille. Käytössä oli perusjako kuuteen päiväryhmään.

#Date	h	real	predict	dif	dif_ %	stDev	pred+2*dev	T_short	T_long	Weekday
30/01/2009	23	8.57	7.58	-0.992	-11.6	0.621	8.82	-9.6	-8.2	Fri
07/01/2009	23	8.52	8.50	-0.016	-0.2	0.621	9.74	-20.4	-16.0	Wed
06/02/2009	23	8.40	8.26	-0.135	-1.6	0.621	9.50	-17.0	-15.9	Fri
16/01/2009	23	8.37	8.92	0.549	6.6	0.621	10.17	-21.3	-21.3	Fri
19/01/2009	23	8.33	7.43	-0.903	-10.8	0.621	8.67	-9.7	-10.1	Mon
08/01/2009	23	8.21	8.51	0.296	3.6	0.621	9.75	-18.2	-18.0	Thu
17/01/2009	23	8.19	8.14	-0.052	-0.6	0.520	9.18	-15.2	-17.5	Sat
18/01/2009	0	8.07	8.02	-0.050	-0.6	0.586	9.20	-12.3	-13.2	Sun
15/01/2009	23	8.07	8.21	0.144	1.8	0.621	9.46	-18.6	-12.5	Thu
07/02/2009	23	8.04	7.99	-0.050	-0.6	0.520	9.03	-16.1	-16.6	Sat
05/02/2009	23	8.03	7.48	-0.545	-6.8	0.621	8.72	-12.0	-8.4	Thu
17/02/2009	23	7.98	9.24	1.251	15.7	0.621	10.48	-25.7	-24.0	Tue
16/02/2009	23	7.96	8.98	1.022	12.9	0.621	10.22	-24.4	-20.9	Mon
17/01/2009	0	7.92	8.81	0.894	11.3	0.663	10.13	-15.2	-17.5	Sat
07/02/2009	0	7.91	8.06	0.146	1.8	0.663	9.38	-16.1	-16.6	Sat

Taulukko 3: Keskimääräisen varaavan sähkölämmittäjän 15 kpl tunteja toteutuneen kulutuksen mukaisessa suuruusjärjestyksessä. Sarakkeessa real on toteutunut kulutus, sarakeessa predict ennusteen antama kulutus ja sarakeessa pred+2\*dev ennuste johon on lisätty analysoitu tunneittainen hajonta kaksinkertaisena (normaalijakaumassa tämä antaa 95 % luottamusatason). Kaikki ennusteen arvot ovat pienempiä kuin  $\mu + 2\sigma$ . T\_short tarkoittaa edellisen 8 tunnin ja T\_long edellisen 40 tunnin lämpötilakeskiarvoa.



Kuva 20: Keskimääräisen öljylämmittäjän tuntikulutukset järjestettynä alkuperäisen mittauksen suuruuden mukaan sinisellä, sekä lilalla ennusteen antamat arvot samoille tunneille. Käytössä oli perusjako kuuteen päiväryhmään.

#Date	h	real	predict	dif	dif %	stDev	pred+2*dev	T_short	T_long	Weekday
14/02/2009	19	1.91	1.53	-0.373	-19.6	0.173	1.88	-11.6	-10.9	Sat
28/02/2009	19	1.82	1.51	-0.300	-16.6	0.173	1.86	-10.8	-10.2	Sat
22/11/2008	18	1.79	1.56	-0.230	-12.8	0.154	1.87	-6.0	-5.6	Sat
21/02/2009	18	1.78	1.50	-0.280	-15.7	0.154	1.81	-7.1	-7.0	Sat
03/01/2009	19	1.74	1.59	-0.154	-8.8	0.173	1.93	-14.4	-16.6	Sat
21/02/2009	19	1.74	1.50	-0.236	-13.6	0.173	1.85	-7.1	-7.0	Sat
31/01/2009	18	1.73	1.61	-0.118	-6.8	0.154	1.92	-17.9	-17.0	Sat
15/11/2008	18	1.71	1.45	-0.262	-15.3	0.154	1.76	0.8	-0.6	Sat
24/01/2009	18	1.70	1.64	-0.063	-3.7	0.154	1.95	-9.9	-8.3	Sat
27/12/2008	18	1.70	1.55	-0.145	-8.5	0.154	1.86	-0.3	0.4	Sat
03/01/2009	18	1.70	1.66	-0.033	-2.0	0.154	1.97	-14.4	-16.6	Sat
09/05/2009	19	1.67	1.25	-0.429	-25.6	0.173	1.59	8.6	7.3	Sat
27/12/2008	17	1.66	1.34	-0.324	-19.5	0.139	1.62	-0.3	0.4	Sat
01/07/2008	0	1.65	0.29	-1.357	-82.3	0.155	0.60	16.9	18.4	Tue
13/12/2008	18	1.63	1.48	-0.148	-9.1	0.154	1.79	-0.3	-1.2	Sat

Taulukko 4: Keskimääräisen öljylämmittäjän 15 kpl tunteja toteutuneen kulutuksen mukaisessa suuruusjärjestyksessä. Sarakkeessa real on toteutunut kulutus, sarakkeessa predict ennusteen antama kulutus ja sarakkeessa pred+2\*dev ennuste johon on lisätty analysoitu tunneittainen hajonta kaksinkertaisena (normaalijakaumassa tämä antaa 95 % luottamusatason). Ne ennusteen antamat arvot, jotka jäivät alle arvon  $\pm 2\sigma$ , on merkitty punaisella. T\_short tarkoittaa edellisen 8 tunnin ja T\_long edellisen 40 tunnin lämpötilakeskiarvoa. 1.7.2008 klo. 0 on mitä luultavimmin mittausvirhe.



Taulukossa 3 on listattuna 15 kappaletta tunteja kuvan 19 vasemmasta laidasta. Näiden tuntien ennustaminen on kaikkein oleellisinta, ja tässä se näyttäisi onnistuneen hyvin. Yksikään mitattu arvo (real) ei ylitä 95 % luottamustasoa ( $\text{pred} + 2 \cdot \text{dev}$ ). Ennusteen arvot nousevat kuitenkin joissain kohdin yllättävänkin korkeiksi. Tähän liittyen varaavan sähkölämmittäjän ennusteessa on joissain tilanteissa ongelmia, joista kerrotaan lisää kappaleessa 7.3.

Kuvassa 20 ja taulukossa 4 on esitettynä edellä esitettyä vastaava informaatio keskimääräiselle öljylämmittäjälle. Öljylämmittäjän ennusteessa virhe näyttää kasvavan kulutuksen kasvaessa, mutta tämä ei välttämättä ole ongelmallinen asia koska jokaiselle tunnille jokaisessa päiväryhmässä lasketaan oma hajonta. Näemme taulukossa 4, että kaikki suurimmat 15 kpl. tuntikulutuksia sattuvat lauantapäivälle kello 18 tai 19.

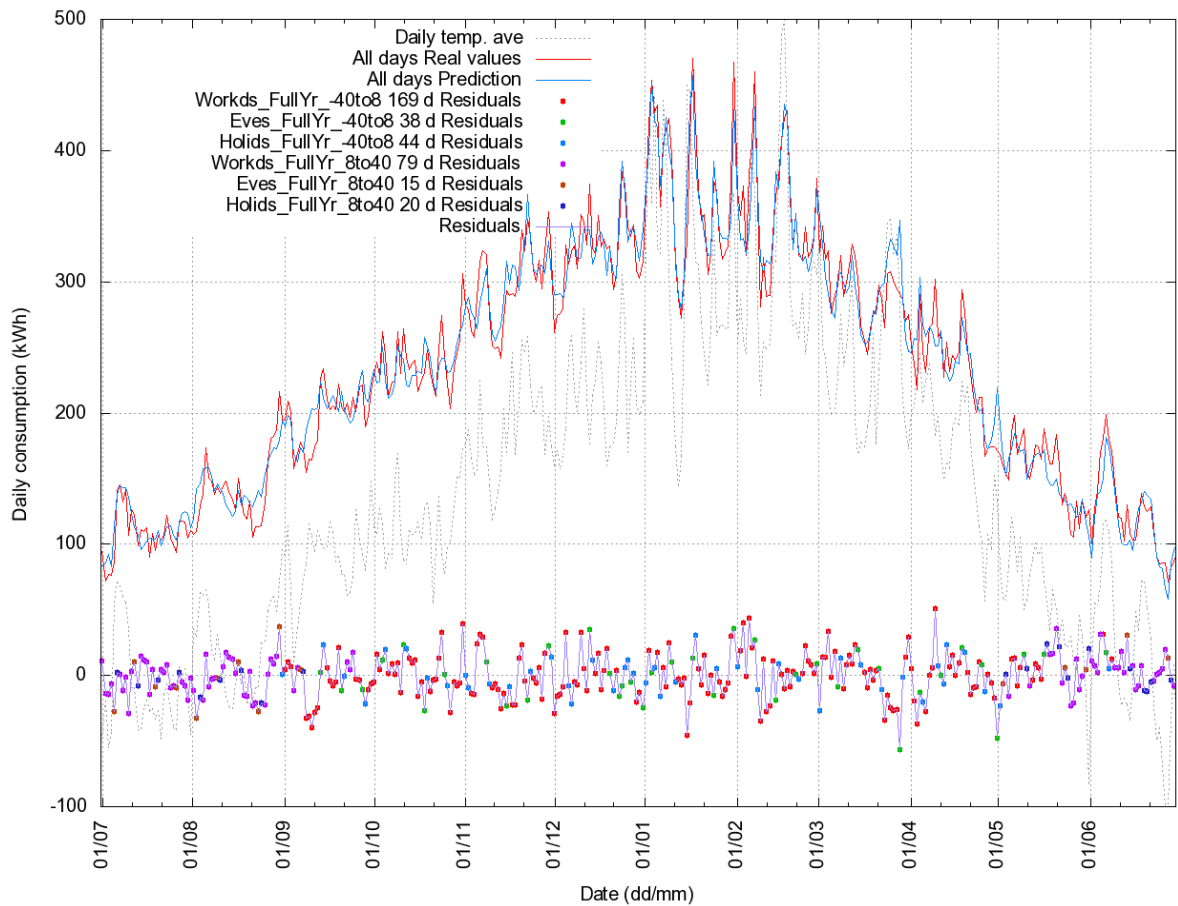
Kuvassa 15 näkyvä päivän korkeimman kulutuksen siirtyminen päivänpituuden pienentyessä kello 20:sta kello 18 pätee myös lauantapäiville. Suurin mitattu tunneittainen kulutus on 1,91 kWh ja ennusteen antama suurin arvo on 1,97 kWh kun lasketaan 95 % luottamustasolla, joten ennustetta voidaan pitää suhteellisen onnistuneena. Tässä esitettyjen keskimääräisen varaavan sähkölämmittäjän ja öljylämmittäjän lisäksi on liitteissä 10 ja 11 esitetty suoran sähkölämmittäjän ja kaukolämpöasiakkaan huippukulutustunnit ja vastaavat pysyvyyskäyrät.

### 6.3 Muuntajan summan ennustaminen

Tätä työtä tehtäessä ei ollut käytettävissä tietoa tietyn muuntajan takana olevista asiakkaista, mutta ohjelman toimintaa on testattu hakemalla sattumanvaraisesti tietty määrä erityyppisiä asiakkaita tietokannasta. Testi on tehty 5, 10, 30 ja 50 asiakkaan ryhmille. Ryhmät voivat koostua erilaisista määristä erityyppisiä asiakkaita, mutta summan päivärakenteen pitäisi olla jonkun edellä käsitellyn lämmitystyyppin kaltainen tai useasta tyypestä koostuva keskiarvo.

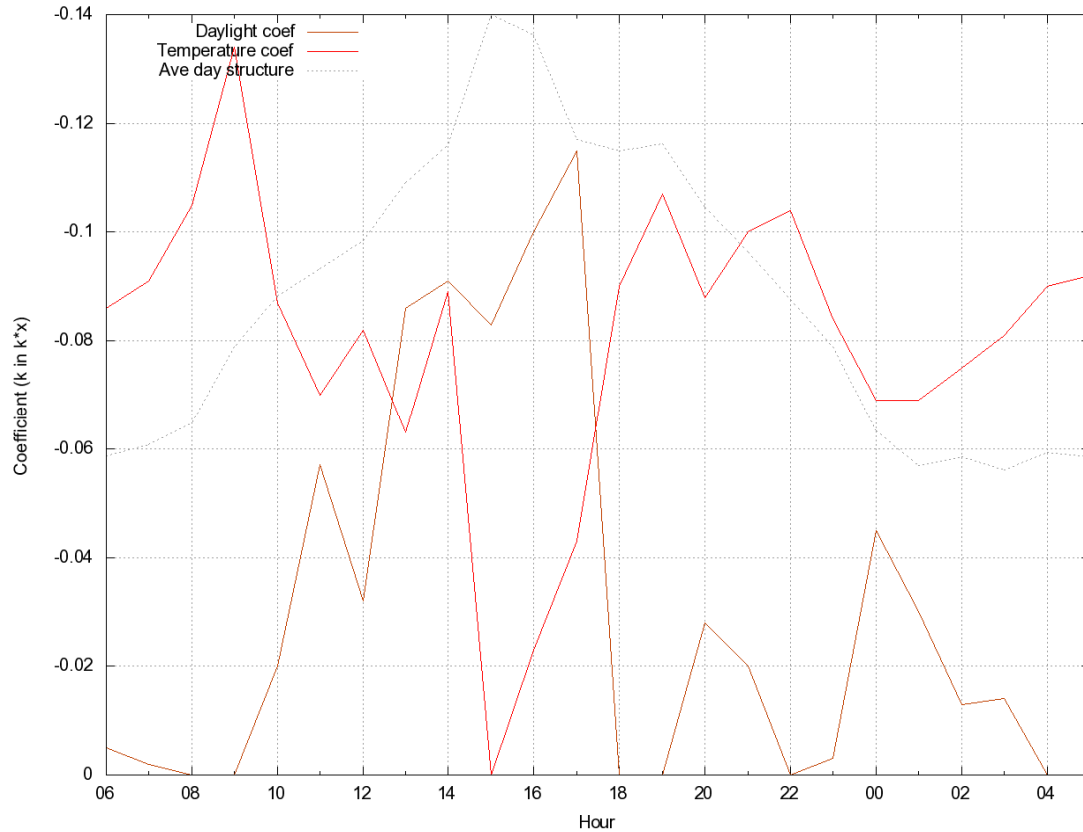
Tässä työssä on arvioitu että asiakkaita täytyy olla vähintään 10 kappaletta jotta löydetään riippuvuus selittävien muuttujien ja sähkönkulutuksen välille. 10 asiakkaan ryhmässäkin on ongelmia, mutta korkean kulutuksen tunnit pystytään yleensä ennustamaan kohtuullisen hyvin. On kuitenkin tärkeää jatkaa testaamista suuremmalla määrällä erikokoisia asiakasryhmiä jotta voidaan antaa luotettavampi arvio minimiasiakasmäärästä. Kuvassa 21 nähdään 10 asiakkaan ryhmän ennuste, joka on onnistunut melko hyvin. Päiväenergioiden mallinnus toimii suhteellisen

hyvin jopa vain 5 asiakkaan ryhmässä, mutta tunneittaisen kulutuksen arviointi ei toimi kovin hyvin kaikissa päiväryhmissä.



Kuva 21: Kymmenen satunnaisesti valitun asiakkaan summan mitattu (punainen) ja ennustettu (sininen) päiväkulutus vuoden aikana, kun vuosi on jaettu kahteen lämpötila-alueeseen ja kolmeen päivätyyppiin. Kuvan alaosassa on kuvattuna ennusteen tekemän virhe eli residuaali vuoden eri päivinä.

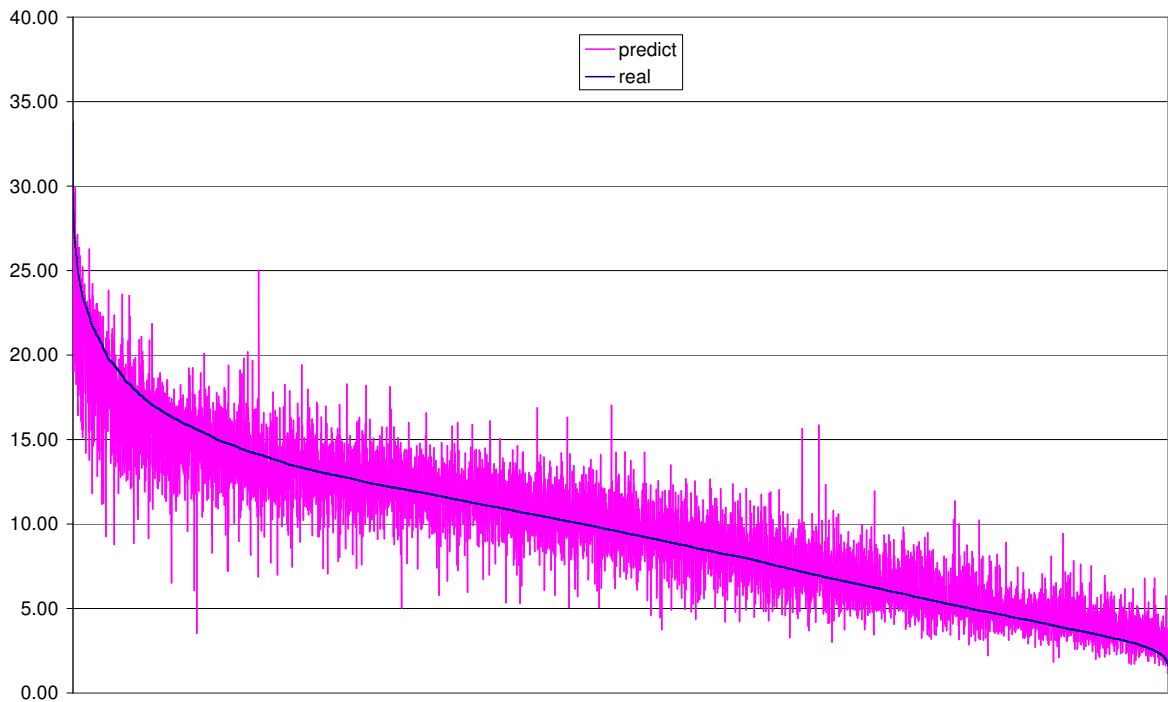
Ongelmia ilmenee niissä ryhmissä joihin ei kuulu kovin montaa päivää. Kesän ryhmissä on ongelmia suurienkin asiakasryhmien kanssa mutta viiden asiakkaan ryhmässä on ongelmia myös kylmien aattopäivien ryhmässä. Tämän huonosti onnistuneen päiväryhmän rakenne näkyy kuvassa 22. Analyysi ei ole pystynyt löytämään jatkuvaa rakennetta päivän sisällä vaan eri tunteja selittävät tekijät muuttuvat arvaamatta, ja välillä on joko päivänpituuden tai lämpötilan vaikutus jouduttu asettamaan nolaksi. Lopputulos ei ole täysin käyttökelpoton mutta suurin piirtein samaan päästäisiin luultavasti vain käyttämällä kullekin tunnille keskiarvoa ja hajontaa. Useammasta kuin viidestä asiakkaista kootut ryhmät saivat aina järkevät päivä rakenteet kylmällä säällä. Kesällä päivä rakenteet eivät ole kovin robusteja, mutta näin voi käydä myös satojen samanlaisten asiakkaiden ryhmissä.



Kuva 22: Esimerkki huonosti onnistuneesta selittävien tekijöiden analyysistä: viiden satunnaisesti valitun asiakkaan summan regressionsovituksen antamat kertoimet aattopäivänä, kun päivän keskiarvolämpötila on alle +8 astetta. Oranssi (päivänpituus) ja punainen (lämpötila) käyrä kuvaavat kertoimia  $b_1$  ja  $b_2$  kaavassa 5.11 eri kellonaikoina. Tässä ne eivät muodosta järkevää päivärakennetta (vrt. järkevät rakenteen kuvissa 16 ja 18). Kerroin asetetaan nollaan jos selittävän muuttujan vaikutus ei täytä kappaleessa 5.4 esitettyjä ehtoja.

#Date	h	real	predict	dif	dif %	stDev	pred+2*dev	T_short	T_long	Weekday
28/12/2008	15	12.72	5.92	-6.799	-53.5	1.476	8.87	-6.1	-3.4	Sun
14/02/2009	18	12.27	5.32	-6.949	-56.6	1.615	8.55	-10.6	-10.2	Sat
24/01/2009	19	11.84	5.74	-6.103	-51.6	1.673	9.08	-12.9	-6.7	Sat
02/01/2009	18	11.80	7.04	-4.761	-40.4	1.533	10.11	-18.2	-11.9	Fri
26/02/2009	18	11.56	5.71	-5.851	-50.6	1.533	8.78	-10.1	-6.6	Thu
23/11/2008	16	11.28	6.03	-5.255	-46.6	1.531	9.09	-3.3	-6.2	Sun
30/01/2009	20	11.11	6.36	-4.755	-42.8	1.532	9.42	-14.8	-6.6	Fri
08/10/2008	17	10.96	4.74	-6.223	-56.8	1.461	7.66	2.3	6.7	Wed
31/12/2008	17	10.89	5.20	-5.691	-52.3	1.498	8.20	-2.4	-1.7	Wed
21/11/2008	17	10.84	5.84	-4.996	-46.1	1.461	8.77	-5.3	-6.0	Fri
31/12/2008	16	10.67	5.84	-4.831	-45.3	1.662	9.16	-2.4	-1.7	Wed
06/03/2009	21	10.59	4.32	-6.271	-59.2	1.415	7.15	-1.8	-1.7	Fri
08/03/2009	20	10.35	5.66	-4.694	-45.4	1.633	8.92	-7.5	-4.0	Sun
22/05/2009	16	10.25	3.70	-6.550	-63.9	1.597	6.89	10.3	11.4	Fri
02/01/2009	17	10.21	6.73	-3.484	-34.1	1.461	9.65	-18.2	-11.9	Fri

Taulukko 5: Tässä on listattuna viiden sattumanvaraisesti valitun asiakkaan 15 kpl tunteja toteutuneen kulutuksen mukaisessa suuruusjärjestyksessä. Sarakkeessa real on toteutunut kulutus, sarakkeessa predict ennusteen antama kulutus ja sarakkeessa pred+2\*dev 95 % luottamusataso. Kaikki 20 suurinta mitattua tuntikulutusarvoa ovat selkeästi suurempia kuin  $\mu+2\sigma$ .



Kuva 23: Kymmenen satunnaisesti valitun asiakkaan tuntikulutukset järjestettynä alkuperäisen mittauksen suuruuden mukaan sinisellä sekä lilalla ennusteen antamat arvot samoille tunneille. Käytössä oli perusjako kuuteen päiväryhmään.

#Date	h	real	predict	dif	dif_ %	stDev	pred+2*dev	T_ short	T_ long	Weekday
07/02/2009	18	33.80	27.56	-6.241	-18.5	4.137	35.83	-14.6	-17.1	Sat
13/12/2008	18	29.57	23.91	-5.658	-19.1	4.137	32.19	0.1	-2.5	Sat
31/01/2009	18	29.30	27.11	-2.188	-7.5	4.137	35.39	-14.4	-14.8	Sat
27/12/2008	18	29.02	22.82	-6.197	-21.4	4.137	31.10	-3.4	0.6	Sat
13/12/2008	19	28.56	19.44	-9.120	-31.9	3.322	26.08	0.1	-2.5	Sat
28/02/2009	18	28.53	23.94	-4.587	-16.1	4.137	32.22	-9.8	-10.5	Sat
17/01/2009	23	27.95	26.38	-1.566	-5.6	1.049	28.48	-14.4	-20.4	Sat
03/01/2009	18	27.81	29.63	1.815	6.5	4.137	37.90	-13.9	-18.2	Sat
27/12/2008	17	27.52	19.01	-8.512	-30.9	3.673	26.35	-3.4	0.6	Sat
18/02/2009	23	27.36	27.39	0.034	0.1	1.279	29.95	-17.9	-24.6	Wed
04/01/2009	23	27.31	26.81	-0.502	-1.8	1.796	30.40	-20.1	-13.9	Sun
06/02/2009	23	27.28	25.67	-1.611	-5.9	1.279	28.23	-17.1	-14.4	Fri
16/01/2009	23	26.94	27.88	0.943	3.5	1.279	30.44	-20.4	-20.8	Fri
24/01/2009	23	26.93	25.63	-1.297	-4.8	1.049	27.73	-12.9	-6.7	Sat
17/02/2009	23	26.92	28.47	1.553	5.8	1.279	31.03	-24.6	-23.7	Tue

Taulukko 6: Kymmenen sattumanvaraisesti valitun asiakkaan 15 kpl tunteja toteutuneen kulutuksen mukaisessa suuruusjärjestyksessä. Sarakkeessa real on toteutunut kulutus, sarakkeessa predict ennusteen antama kulutus ja sarakkeessa pred+2\*dev ennuste johon on lisätty analysoitu tunneittainen hajonta kaksinkertaisena (normaalijakaumassa tämä antaa 95 % luottamusatason). Ne ennusteen antamat arvot, jotka jäivät alle arvon  $\mu+2\sigma$ , on merkitty punaisella. T\_short tarkoittaa edellisen 8 tunnin ja T\_long edellisen 40 tunnin lämpötilakeskiarvoa.

Taulukossa 5 nähdään viiden satunnaisesti valitun asiakkaan summan 15 huippukulutustuntia. Ennuste antaa aivan liian matalia arvoja, eikä hajonnan huomioiminen auta asiaa. Toteutunut huippukulutus on yli kaksi kilowattituntia suurempi kuin ennustettu. Viiden asiakkaan ryhmä näyttäisi olevan liian pieni kunnollisen ennusteen aikaansaamiseksi, mutta yleensä kymmenen asiakkaan ryhmälle ennuste onnistuu jo paljon paremmin.

Kymmenenkin asiakkaan ryhmän ennuste saattaa kuitenkin epäonnistua jos asiakkaat sattuvat olemaan vaikeasti ennustettavia. Kuvassa 23 näemme 10 asiakkaan tunneittaisen pysyvyyskäyrän. Virheet näyttävät melko suurilta, mutta taulukosta 6 huomaamme että hajonnan huomioimisen jälkeen 95 % luottamustaso kattaa suurimman osan mitatuista arvoista.

Taulukossa 6 nähdään ero eri päivätyyppien eri tunneissa. Huipputunnit osuvat yleensä lauantaille klo 18 tai satunnaisille päivillä klo 23. Lauantain huipun voi selittää esimerkiksi saunan käyttö, ja koska joinain lauantainpäivinä saattaa lähes kaikilla olla sauna päällä klo. 18 ja toisina päivinä ei kenelläkään, saadaan suuri hajonta: tässä yli neljä kilowattituntia. Yösähkön käytön aiheuttama klo 23:n piikki taas ei muutu yhtä voimakkaasti, joten hajonta on paljon pienempi: tässä hieman yli yksi kilowattitunti. Kymmenen asiakkaan ryhmän ennuste tuntuu melko onnistuneelta, mutta tarvitaan lisää testausta suuren määrän noin kymmenen asiakkaan ryhmillä jotta voidaan antaa suositus minimiasiakasryhmäkoosta.

## 7 Ennuste vuoden 2010 tammikuulle

Aikaisemmin tässä työssä on käsitelty lähtödatan ja siitä tehdyn ennusteen välistä eroa. Työn loppuvaiheessa Kainuusta saatiin kuitenkin testidataksi myös vuoden 2010 tammikuun kulutusdata. Tämä data sisälsi suurimman osan niistä asiakkaista, joilta alkuperäinen data saatiin. Myös ilmatieteenlaitokselta saatiin lämpötiladata Kainuun lentokentältä samalta ajanjaksolta.

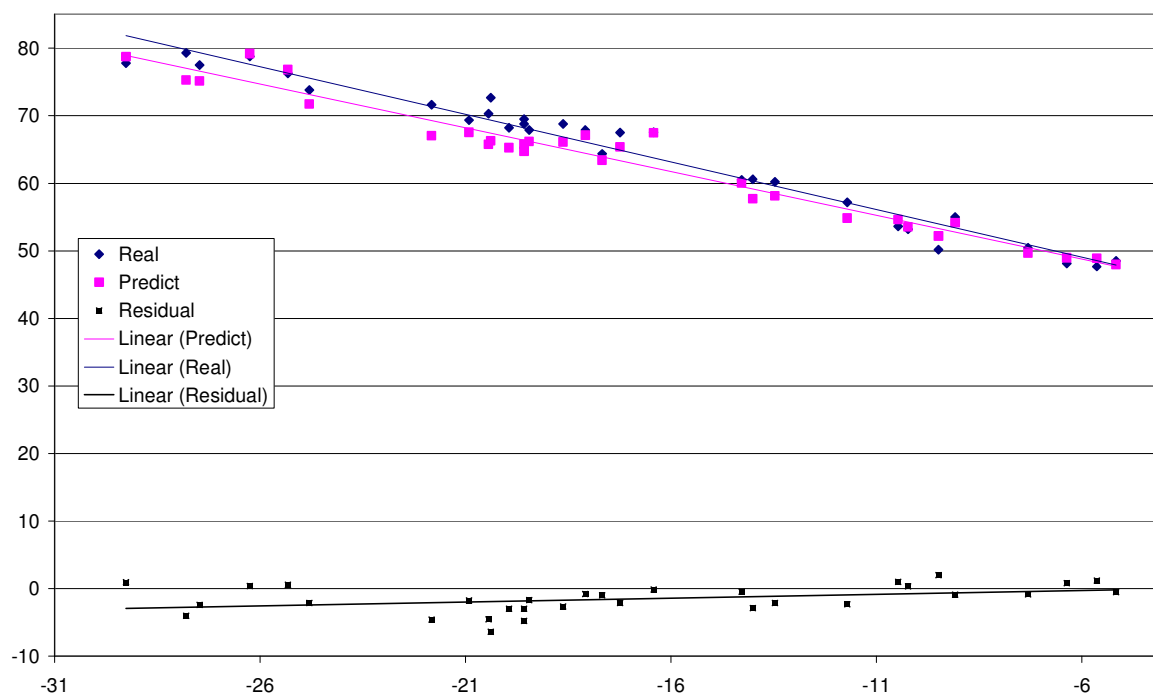
Ennusteen lähtödatana käytetyn vuoden 2008–2009 (1.7.2008–30.6.2009) talvi oli melko lämmin, kun taas vuoden 2010 tammikuu oli selvästi kylmempi. Tammikuun 2010 tunneittaisten mittausten keskilämpötila Kainuun lentokentällä oli  $-17,0$  astetta, kun taas vuoden 2009 tammikuussa se oli  $-10,2$  astetta. 2010 tammikuun matalin lämpötila oli  $-34,1$  astetta, ja 2009 -  $-26,8$  astetta. Kuvasta 5 voidaan arvioida lineaarinen kulutuksen nousu suoralle sähkölämmittäjälle: noin  $1,5$  kWh/aste päiväkulutuksissa. Varaavalle sähkölämmittäjälle se on päiväkulutuksissa samaa luokkaa.

Suoran sähkölämmittäjän keskimääräinen päiväkulutus 01/2009 oli  $67,8$  kWh. Suuruusluokka etsittäessä voidaan verrata keskimääräistä lämpötilaeroa 2009 ja 2010 tammikuun välillä: tämä oli  $n. 7$  astetta. Oletettu päiväkulutuksen nousu on siis 01/2010 sähköä lämmitykseen käyttävillä asiakkailla  $(7 \cdot 1,5) / 67,8 = 15,5 \%$ . Tämä on hyvin karkea arvio. mutta sitä voidaan käyttää suuruusluokka-arviona.

Melko suuri lämpötila ero 2009 ja 2010 tammikuun välillä on testiä varten hyvä asetelma, koska oletuksena sähkön kulutukseltaan pienemmän aikajakson avulla ennustetaan oletuksena korkeampaa kulutusta. Tässä näemme kuinka hyvin ohjelma pystyy huomioimaan lämpötilan muutoksen, ja toimiiko lineaarinen oletus lämpötilan ja kulutuksen suhteesta käytännössä.

### 7.1 Eri lämmitystyyppit tammikuussa 2010

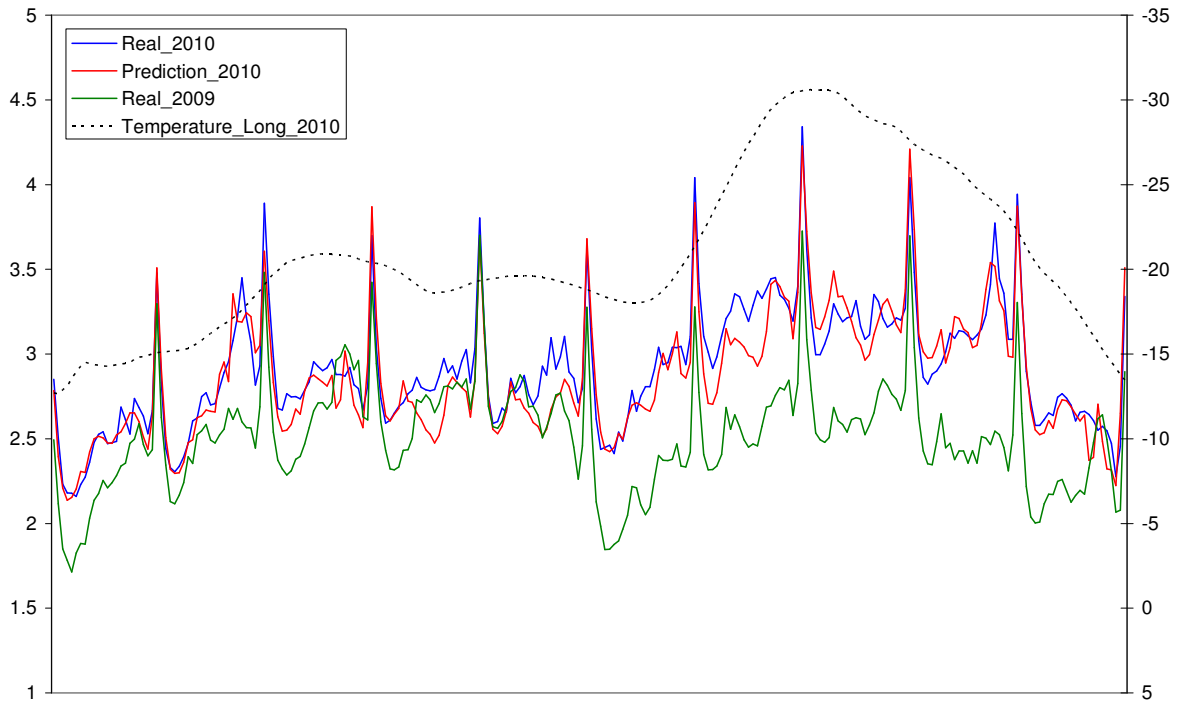
Tammikuun 2010 kulutusennuste testattiin erityyppisillä asiakasryhmillä. Suoraa sähkölämmitystä käyttävien asiakkaiden ennuste onnistui hyvin vaikka aivan kaikilta niistä asiakkaista joilta saatiin vuoden 2008–2009 kulutusdata, ei saatu tammikuun 2010 dataa. Koska testissä on asiakkaiden keskiarvo, ei muutaman asiakkaan puuttuminen haittaa.



Kuva 24: Keskimääräisen suoran sähkölämmittäjän päiväkulutukset (kWh) tammikuussa 2010, kun x-akselilla on vuorokauden keskiarvolämpötila. Tumman sinisellä on merkitty toteutuneet kulutukset 2010 ja lilalla 1.7.2008–30.6.2009 datalla tehdyn ennusteen arvio. Mustalla on merkitty näiden välinen ero eli tehty virhe. Päivät oli jaettu kahteen lämpötila-alueeseen kohdasta +8 astetta ja kolmeen päivätyyppiin

Hieman erikoinen asia oli usean yksittäisen asiakkaan melko suuri ero 2009 ja 2010 tammikuun päiväkeskiarvokulutuksen välillä. Joillakin asiakkailla kulutus oli pienempi ja toisilla jopa kaksi kertaa suurempi. Asiaa tutkittaessa huomattiin että asiakkailta saatu mittausdata oli kuitenkin järkevän näköistä eikä ongelmia mittauksissa ollut. Keskiarvo käyttäytyi todellisuudessa melko hyvin ennusteen mukaisesti. Keskimäärin tammikuun päiväkulutus oli vuonna 2010 suoraa sähkölämmitystä käyttävillä asiakkailla 18,9 % suurempi kuin vuonna 2009, eli ero oli huomattava.

Kuvassa 24 nähdään että suoran sähkölämmittäjän ennusteesta saadaan jonkin verran liian alhainen päiväenergiaennuste, ja keskimäärin kylmemmällä säällä virhe on suurempi. Kuitenkin kylmimpinä päivinä ennuste osuu melko hyvin kohdalleen. Ylipäätään kulutus on hieman arvioitu suurempi, mutta ei kohtuuttoman paljon. Ennuste on tehty käyttämällä tässä työssä yleisesti käytettyä kahden lämpötila-alueen jakoa, jolloin suurin osa kevät-, syksy- ja talvipäivistä kuuluu samaan päiväryhmään.



Kuva 25: Keskimääräisen suoran sähkölämmittäjän tunneittainen mitattu (sinisellä) sekä ennustettu kulutus (punaisella) ajalta 1-10.12.2010. Musta katkoviiva kuvaa edellisen 40 tunnin lämpötilakeskiarvoa, ja vihreällä katkoviivalla on kuvaan merkitty vuoden 2009 kulutus samalta ajalta. Päivät oli jaettu kahteen lämpötila-alueeseen kohdasta +8 astetta ja kolmeen päivätyyppiin.

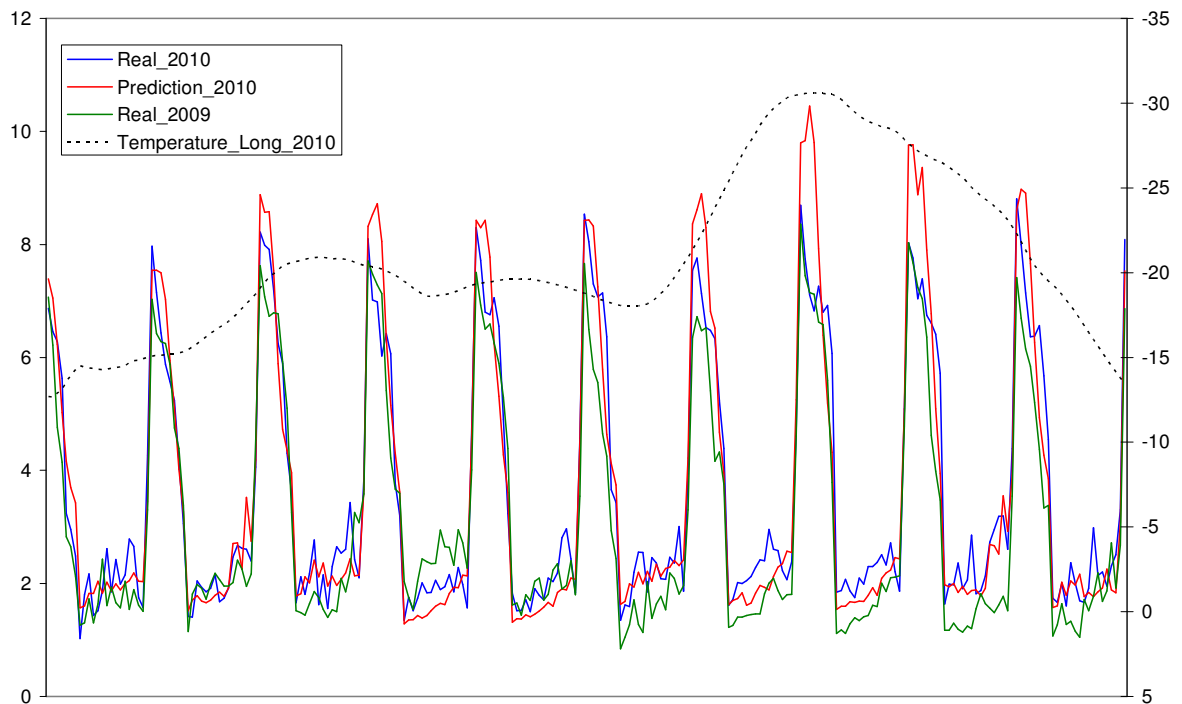
#Date	h	real	predict	dif	dif_ %	stDev	pred+2*dev	T_short	T_long	Weekday
07/01/2010	23	4.34	4.23	-0.112	-2.6	0.094	4.42	-29.7	-30.5	Thu
23/01/2010	23	4.11	4.06	-0.056	-1.4	0.108	4.27	-28.0	-27.7	Sat
22/01/2010	23	4.08	4.01	-0.073	-1.8	0.094	4.20	-25.7	-26.0	Fri
28/01/2010	23	4.05	3.76	-0.288	-7.1	0.094	3.95	-22.7	-21.7	Thu
08/01/2010	23	4.04	4.21	0.168	4.2	0.094	4.40	-26.9	-27.6	Fri
06/01/2010	23	4.04	3.89	-0.146	-3.6	0.144	4.18	-25.7	-21.3	Wed
21/01/2010	23	3.97	3.82	-0.151	-3.8	0.094	4.01	-24.2	-23.5	Thu
24/01/2010	23	3.96	4.17	0.206	5.2	0.144	4.46	-19.9	-24.6	Sun
09/01/2010	23	3.94	3.87	-0.071	-1.8	0.108	4.09	-19.7	-22.3	Sat
26/01/2010	23	3.91	3.69	-0.220	-5.6	0.094	3.88	-20.1	-20.7	Tue
02/01/2010	23	3.89	3.61	-0.285	-7.3	0.108	3.82	-21.7	-19.1	Sat
27/01/2010	23	3.84	3.69	-0.154	-4.0	0.094	3.88	-18.6	-19.6	Wed
29/01/2010	23	3.84	3.73	-0.109	-2.8	0.094	3.92	-18.2	-19.9	Fri
23/01/2010	18	3.82	3.62	-0.195	-5.1	0.151	3.93	-27.4	-26.7	Sat
04/01/2010	23	3.80	3.68	-0.123	-3.2	0.094	3.87	-19.7	-19.3	Mon

Taulukko 7: Keskimääräisen suoran sähkölämmittäjän vuoden 2010 tammikuun 15 kpl suurinta tuntikulutusta. Sarakkeessa real on toteutunut kulutus 2010, sarakeessa predict ennusteen antama kulutus ja sarakeessa pred+2\*dev ennuste johon on lisätty analysoitu tunneittainen vuoden 2009 hajonta kaksinkertaisena (normaalijakaumassa tämä antaa 95 % luottamusatason). Punaisella merkityt toteutuneet kulutukset ylittivät hieman arvon  $\mu+2\sigma$ . T\_short tarkoittaa edellisen 8 tunnin ja T\_long edellisen 40 tunnin lämpötilakeskiarvoa vuonna 2010.



Vuoden 2010 tammikuun alun toteutunut kulutus sekä analyysin antama ennuste suoralla sähkölämmittäjälle nähdään kuvassa 25. Kulutus vaihtelee voimakkaasti lämpötilan mukaan, joten ennusteen tekeminen oli helppoa. Lämpötilan muutoksien huomioon ottamisen lisäksi aatto- ja pyhäpäivien erottaminen on tärkeä osa ennusteen onnistumista. Taulukossa 7 on listattuna tammikuun 2010 huippukulutustunnit: ennuste onnistui hyvin ja oletetusti kello 23 kulutus oli korkein.

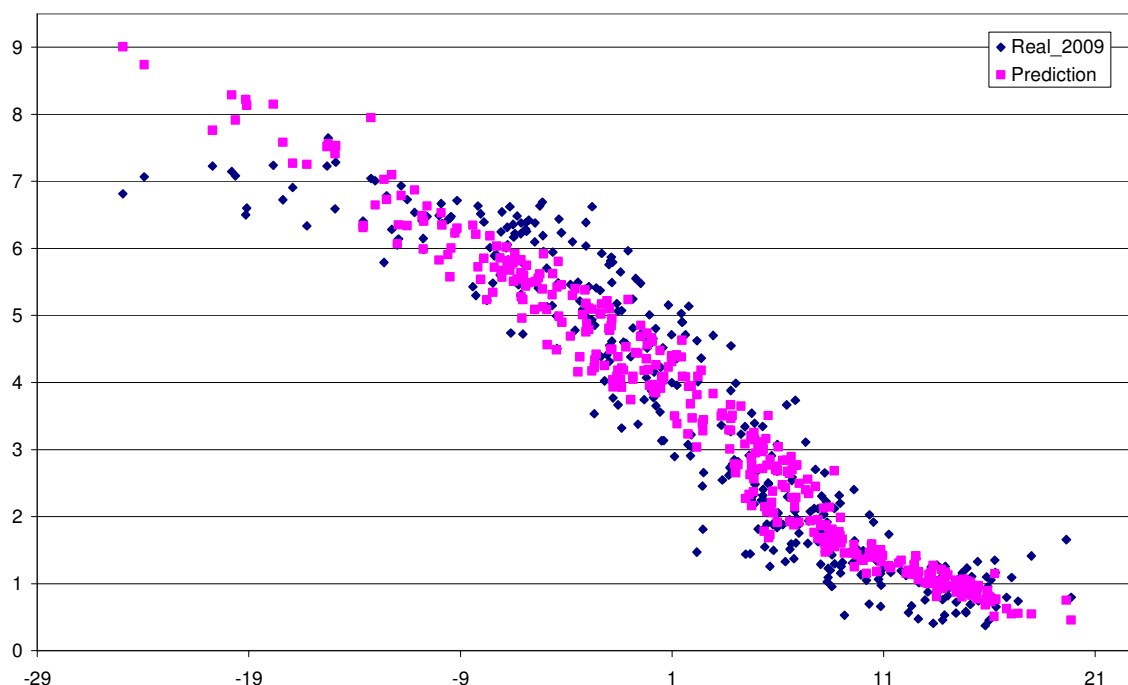
Varaavan sähkölämmityksen ennuste onnistui päiväenergioita tarkasteltaessa paremmin kuin edellä käsitellyn suoran sähkölämmityksen ennuste. Tätä edelsi kuitenkin yhden asiakkaan poisto tutkittavasta datasta. Vain 21:llä varaavaa sähkölämmitystä käyttävältä asiakkaalta on dataa sekä 2009 että 2010 tammikuulta. Näistä kaikkein suurimman asiakkaan keskimääräinen päiväkulutus oli pudonnut vuoden 2010 tammikuussa vain 13 %:iin siitä, mitä se oli ollut tammikuussa 2009. Jos asiakas olisi ollut mukana keskiarvoa määritettäessä, olisi vuoden 2010 tammikuun kulutus ollut vain 6 % prosenttia suurempi kuin 2009.



Kuva 26: Keskimääräisen varaavan sähkölämmittäjän tunneittainen mitattu (sinisellä) ja ennustettu kulutus (punaisella) ajalta 1-10.12.2010. Musta katkoviiva kuvaa edellisen 40 tunnin lämpötilakeskiarvoa, ja vihreällä katkoviivalla on kuvaan merkitty vuoden 2009 kulutus samalta ajalta. Päivät oli jaettu kahteen lämpötila-alueeseen kohdasta +8 astetta ja kolmeen päivätyyppiin.

Verrattuna suoran sähkölämmityksen 18,9 % nousuun, 6 % on erittäin pieni luku. Kun mainittu yksittäinen asiakas poistetaan sekä vuoden 2009 että 2010 datasta, saadaan 01/2010 päiväkeskiarvoksi 16,3 % suurempi arvo kuin 01/2009. Tämä tuntuu sähköä lämmitykseen käyttävien asiakkaiden tapauksessa paljon järkevämmältä arvolta, joten tätä yhden asiakkaan vajaata dataa käytettiin testauksessa.

Tunneittaista kulutusta ennustettaessa varaavan sähkölämmityksen kanssa on kuitenkin suurempia ongelmia. Kuvisa 17 ja 18 nähdään 2008–2009 datan avulla tehdyn ennusteen mukainen varaavan sähkölämmittäjän päivärakenne ja selittävien muuttujien kertoimet eri tunteina. Lämpötilan vaikutuskerroin on tunteilla 1 ja 2 selvästi suurin. Ennusteen mukainen kulutushuippu kuitenkin sattuu tunteille 1 ja 2 vain erittäin kylminä päivinä, koska kello 23 kulutus on tasaisesti erittäin korkea.



Kuva 27: Keskimääräisen varaavan sähkölämmittäjän kello yhden mitattu tuntikulutus ajalta 1.7.2008–30.6.2009 (siniset pisteet) ja siitä tehty ennuste (lilat pisteet). X-akseliksi on valittu 40 tunnin lämpötilakeskiarvo. Ennuste on tehty kahdella lämpötila-alueella, jonka rajaa lämpötila +8 astetta.

Tammikuussa 2010 oli kylmää, ja edellisen vuoden mukaan tehty ennuste antaa liian suuria arvoja varsinkin tunteille 1 ja 2. Tämä ongelma nähdään selvästi kuvassa 26. Kuvan 18 mukaisesti tuntien 1 ja 2 lämpötilavaikutuskerroin on erittäin suuri, mutta se on saatu arvioimalla lineaarinen

riippuvuus suuressa ryhmässä päiviä. Oikeasti tunneilla 1 ja 2 riippuvuus lämpötilan ja sähkön kulutuksen välillä ei kuitenkaan ole lineaarinen, vaan varaavaa sähkölämmitystä käyttävillä asiakkailla on tietty tehomäärä jota enempää asiakas ei ota tehoa verkosta. Tämä taas muuttaa tuntien 3 ja 4 kulusta korkeammaksi koska vaadittava lämmitysenergia täytyy saada kokoon.

Varaavan sähkölämmittäjän ei-lineaarinen käyttäytyminen nähdään osittain jo vuoden 2009 datassa, mutta virheet kasvavat suuriksi vasta ennustettaessa vuoden 2010 kylmän tammikuun kulutusta. Kuvassa 27 on esitetty kahden lämpötila-alueen avulla tehty ennuste ja mitatut vuoden 2008–2009 arvot tunnille 1. Lopputulos olisi parempi käytettäessä kolmea lämpötila-aluetta, mutta hyvin kylmien päivien ennustaminen on joka tapauksessa vaikeaa koska niitä on kovin vähän. Varaavan sähkölämmityksen erityisluonnetta ja sen hallitsemisesta ennusteesta käsitellään tarkemmin kappaleessa 7.3. Esimerkkijakso ja huippukulutustuntitaulukko keskimääräisen kaukolämpöasiakkaan ja öljylämmittäjän ennusteesta ajalta 01/2010 nähdään liitteissä 12 ja 13.

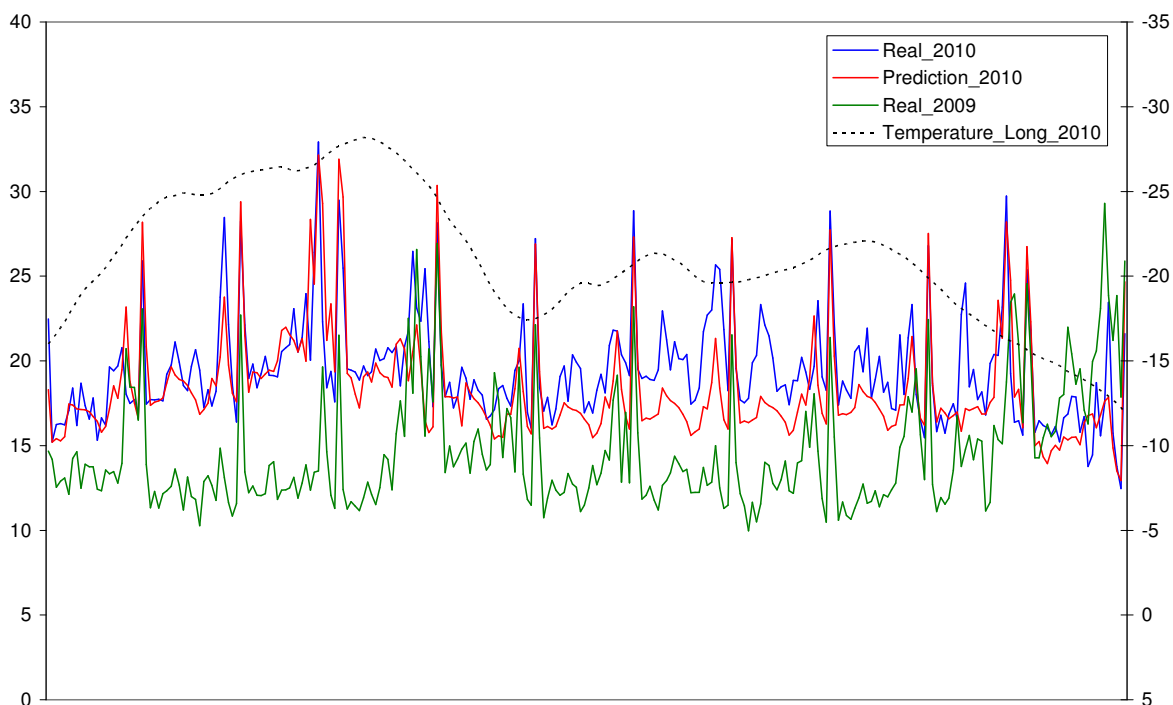
## 7.2 Muuntajan summakuormitus

Muuntajien summakuormitusta käsitellään tässä samalla tavalla kuin aiemmin tukittaessa vuoden 2008–2009 dataa. Testiin on valittu 5, 10 ja 30 asiakkaan sattumanvaraisesti valittu ryhmä. Näistä viiden asiakkaan ryhmän ennuste ei onnistunut kovin hyvin, mutta tämä oli oletettavaa jo vuoden 2008–2009 datasta muodostetun ennusteen analyysin perusteella (kappale 6.3 ). Kuitenkin jo 10 asiakkaan ryhmälle saatiin suhteellisen hyvältä näyttävä ennuste, kuten nähdään kuvassa 28.

Taulukossa 8 on esitetty 10 asiakkaan summan huippukulutustunnit 01/2010 ja niiden ennuste. Ainoastaan yksi lauantain kello 16 tunti on ollut hajonnan analyysin jälkeen yllättävän korkea: tämä on kuitenkin ennustettavaa, koska lauantai-iltapäivän ja -illan kulutuksella on suuri hajonta ja yllättäviä kulutushuippuja esiintyy useammin kuin muilla tunneilla.

Pienien asiakasryhmien ennustamista tulisi testata enemmän, jotta voidaan arvioida paremmin mikä on riittävä määrä asiakkaita ennusteen onnistumiseksi. Tämä tietenkin riippuu asiakastyypeistä. Olisikin hyvä saada testiin todellisia muuntajien takana olevia ryhmiä, koska kaupunkialueella yksi muuntaja syöttää yleensä useita asiakkaita kun taas maaseudulla muuntaja voi syöttää vain yhtä asiakasta. Näitä yhden tai vain muutaman asiakkaan ryhmiä tulisi testata jotta voidaan arvioida onko mahdollista luotettavasti ennusteen jakelumuuntajatasen kulutus

kaikissa tilanteissa vai onko asiakkaiden määrä rajoittava tekijä jolloin ei voida määrittää ennusteen riittävyyttä millekään tietylle verkon topologiselle tasolle.



Kuva 28: Kymmenen satunnaisesti valitun asiakkaan tunneittainen mitattu kulutus (sinisellä) sekä ennustettu kulutus (punaisella) ajalta 21–31.12.2010. Musta katkoviiva kuvaa edellisen 40 tunnin lämpötilakeskiarvoa, ja vihreällä katkoviivalla on kuvaan merkitty vuoden 2009 kulutus samalta ajalta. Päivät oli jaettu kahteen lämpötila-alueeseen kohdasta +8 astetta ja kolmeen päivätyyppiin.

#Date	h	real	predict	dif	dif_%	stDev	pred+2*dev	T_short	T_long	Weekday
23/01/2010	18	32.92	32.15	-0.769	-2.3	4.234	40.62	-27.4	-26.7	Sat
07/01/2010	19	30.40	25.25	-5.153	-17.0	2.864	30.97	-31.3	-30.1	Thu
30/01/2010	18	29.73	28.20	-1.531	-5.1	4.234	36.67	-15.2	-16.3	Sat
23/01/2010	23	29.48	31.90	2.418	8.2	0.978	33.85	-28.0	-27.7	Sat
08/01/2010	23	29.35	30.75	1.397	4.8	1.271	33.29	-26.9	-27.6	Fri
03/01/2010	23	29.33	28.44	-0.895	-3.1	1.802	32.04	-18.0	-20.4	Sun
04/01/2010	23	29.04	27.39	-1.651	-5.7	1.271	29.93	-19.7	-19.3	Mon
09/01/2010	18	28.98	32.10	3.120	10.8	4.234	40.57	-21.6	-23.9	Sat
26/01/2010	23	28.86	27.31	-1.546	-5.4	1.271	29.85	-20.1	-20.7	Tue
28/01/2010	23	28.83	27.72	-1.106	-3.8	1.271	30.26	-22.7	-21.7	Thu
22/01/2010	19	28.46	23.76	-4.698	-16.5	2.864	29.49	-25.4	-25.4	Fri
16/01/2010	16	28.29	20.35	-7.941	-28.1	3.025	26.40	-6.6	-8.4	Sat
22/01/2010	23	28.24	29.38	1.141	4.0	1.271	31.92	-25.7	-26.0	Fri
24/01/2010	23	28.14	30.36	2.224	7.9	1.802	33.97	-19.9	-24.6	Sun
07/01/2010	23	28.07	30.86	2.793	9.9	1.271	33.40	-29.7	-30.5	Thu

Taulukko 8: Kymmenen satunnaisesti valitun asiakkaan vuoden 2010 tammikuun 15 kpl suurinta tuntikulutusta. Sarakkeessa real on toteutunut kulutus 2010, sarakkeessa predict ennusteen antama kulutus ja sarakkeessa pred+2\*dev 95 % luottamusarvonnuste. Punaisella merkityt toteutuneet kulutukset ylittivät arvon  $\pm 2\sigma$ . T\_short tarkoittaa edellisen 8 ja T\_long 40 tunnin lämpötilakeskiarvoa vuonna 2010.

### 7.3 Varaavan sähkölämmittäjän kulutusennuste

Varaavan sähkölämmittäjän päivärakenne on haasteellinen, koska päivän tunnit käyttäytyvät hyvin eri lailla. Päivällä sähköä käytetään vähän ja yöllä halvemman sähkön aikaan sitä käytetään paljon. Tässä työssä esitetyllä laskennalla saadaan kohtuullisesti käsiteltyä eri tuntien kulutuksen vaihtelu lämpötilan sekä päivänpituuden mukaisesti.

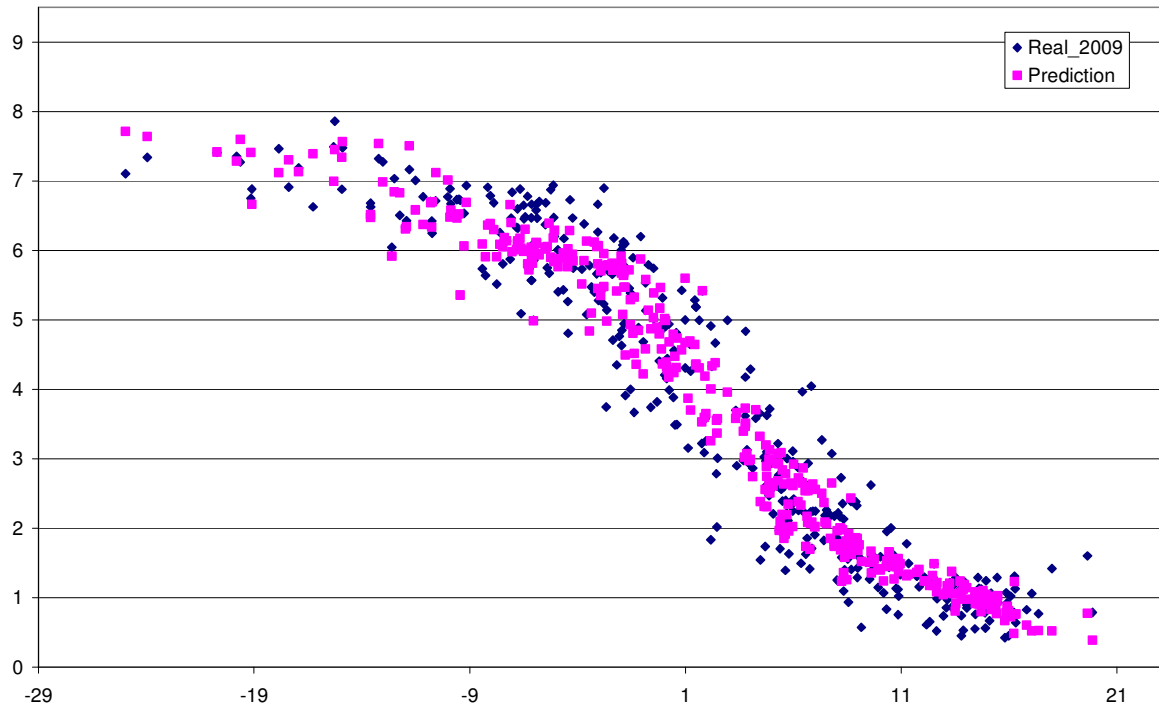
Kappaleessa 7.1 kuitenkin havaittiin ongelmaksi tiettyjen tuntien, lähinnä tuntien 1 ja 2, epälineaarisuus hyvin kylminä päivinä. Tässä työssä lämpötilan vaikutus kulutukseen oletetaan lineaariseksi, ja epälineaarisuus otetaan huomioon ainoastaan jakamalla päivät lämpötilaryhmiin. Yleisesti tässä työssä on päivät jaettu kahteen alueeseen kohdasta +8 astetta (päivän keskiarvolämpötila). Tämä toimii lähes kaikille asiakastyypeille ja päivän eri tunneille kohtuullisen hyvin, ja takaa riittävän määrän päiviä kuhunkin ryhmään.

Varaavan sähkölämmittäjän tapauksessa näyttäisi 2010 tammikuun testin mukaan olevan kuitenkin tarvetta jakaa kylmät päivät vielä kahteen lämpötilaryhmään tai vaihtoehtoisesti käyttää jotain muuta kuin lineaarista yhteyttä lämpötilan ja kulutuksen välillä. Tässä on päätetty käyttää lineaarista sovitusta koska se takaa matemaattisesti vakaan ja helpon käsittelyn, ja toisaalta koska tämän muuttaminen veisi niin paljon aikaa että se jätetään tämän työn ulkopuolelle.

Fysikaaliselta kannalta tarkasteltuna ei-lineaarinen sovitus ei välttämättä ole kovin järkevä: kesällä ei-lineaarisuus johtuu siitä että lämmitys lopetetaan kokonaan ja talvella luultavasti siitä että tiettyä maksimia suurempaa lämmitystehoa ei saada varaavaan lämmitykseen syötettyä. Kun tietty maksimiteho saavutetaan, siirtyy lämmitysenergiatarve seuraaville tunneille. Yhtä tuntia, esimerkiksi kello 1, käsiteltäessä tämä näkyy, ei niinkään epälineaarisuutena, vaan epäjatkuvuutena kahdessa kohtaa: kesällä kun lämmitys loppuu ja talvella kun lämmitystehohuippu saavutetaan. Tämä näkyy myös kuvassa 27 vaikka siinä muutos ei olekaan täysin yhtäkkinen. Yhtäkkisyys olisi tässä hyvin epätodennäköistä koska kyseessä on usean asiakkaan keskiarvo.

Varaavan sähkölämmityksen tapauksessa eri tunnit vaikuttavat toisiinsa, joten jos esimerkiksi tunnilla 1 saavutetaan maksimiteho, siirtyy lämmitystarve tunnille 2 jne. Tämän takia tarvittaisiin melko monimutkainen analyysi, jossa olisi joka tapauksessa tiedettävä maksimiteho. Maksimiteho ei kuitenkaan ole vakio kuvaa 27 ja taulukko 9 vertailtaessa. Kuvasta 27 voi arvioida että teho ei mene yli 8 kW:iin, mutta toisaalta taulukossa 9 nähdään että yhden tunnin mitattu kulutus,

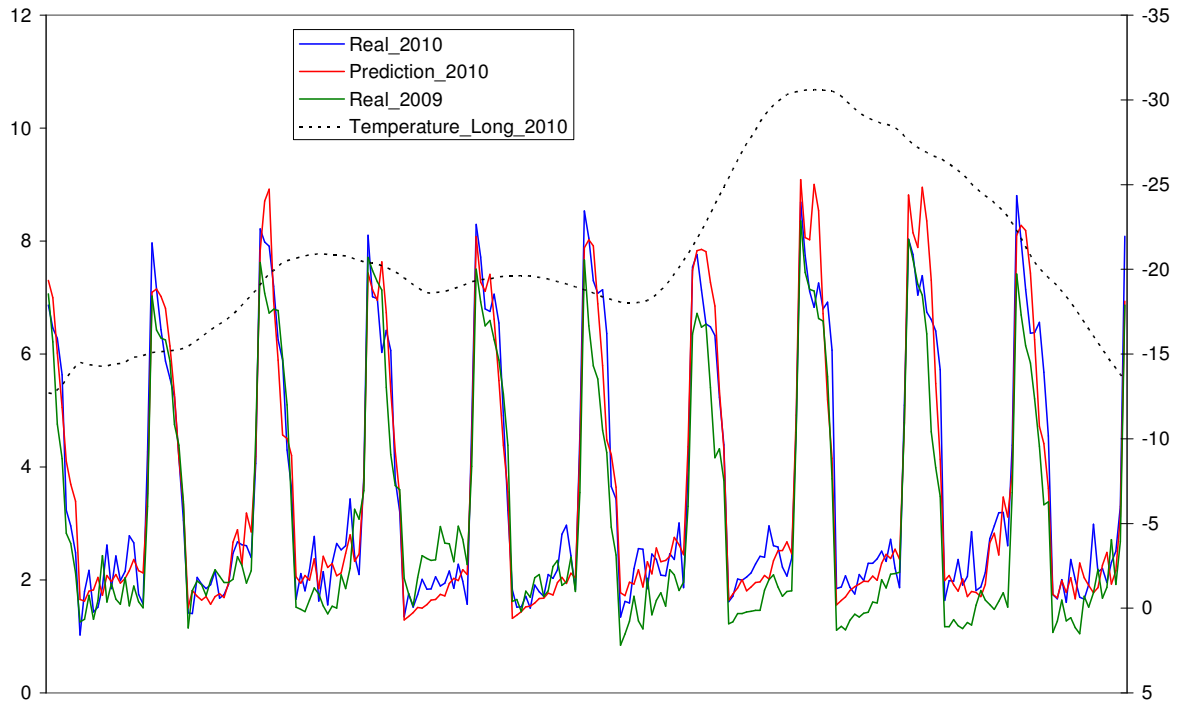
vaikkakin tunnin 23, oli ollut yli 9 kW. Tässä työssä halutaan analyysi pitää nopeana ja selkeänä joten usean tunnin vaikutusta toisiinsa ei lähdetty tutkimaan.



Kuva 29: Keskimääräisen varaavan sähkölämmittäjän kello yhden mitattu tuntikulutus ajalta 1.7.2008–30.6.2009 (siniset pisteet) ja siitä tehty ennuste (lilat pisteet). X-akseliksi on valittu 40 tunnin lämpötilakeskiarvo. Ennuste on tehty kolmella lämpötila-alueella, joita rajaavat lämpötilat -2 ja +8 astetta.

Nopea ratkaisu, joka tuo tilanteeseen selkeän parannuksen, on käyttää kolmea lämpötila-aluetta. Tästä aiheutuu kuitenkin jo aiemmin huomioitu ongelma, eli yksittäiseen päiväryhmään jää vähemmän päiviä. Koska päivät jaetaan erikseen aatto-, pyhä- ja arkipäiviin, jää joihinkin ryhmiin hyvin vähän päiviä. Jos ryhmässä on paljon asiakkaita, mielellään useita kymmeniä, saadaan silloinkin yleensä aikaan melko hyvä ennuste. Kuvassa 29 nähdään kuvaa 27 vastaava kello 1 käyttäytyminen kun käytössä on kolme lämpötila-aluetta: ennuste on huomattavasti parempi.

Kuvassa 30 nähdään ennuste tammikuun alussa 2010 kun käytetään kolmea lämpötila-aluetta. Kylmemmät päivät jaetaan kohdasta -2 astetta, mikä on melko korkea raja mutta varmistaa että kaikkiin ryhmiin saadaan kohtuullinen määrä päiviä. Tällä tavalla lopputulos on huomattavasti parempi kuin kuvassa 26 kahdella alueella. Kuviin 30 ja 26 on valittu ennusteen huonoin sopivuus tammikuussa 2010 jotta virheet tulevat esille. Niissä edellisen vuoden arvot olisivat suoraan ennustaneet kulutuksen paremmin kuin ohjelmalla tehty ennuste. Tilanne ei kuitenkaan ole ennusteen kannalta niin huono muussa osassa tammikuuta 2010, kuten nähdään liitteessä 14.



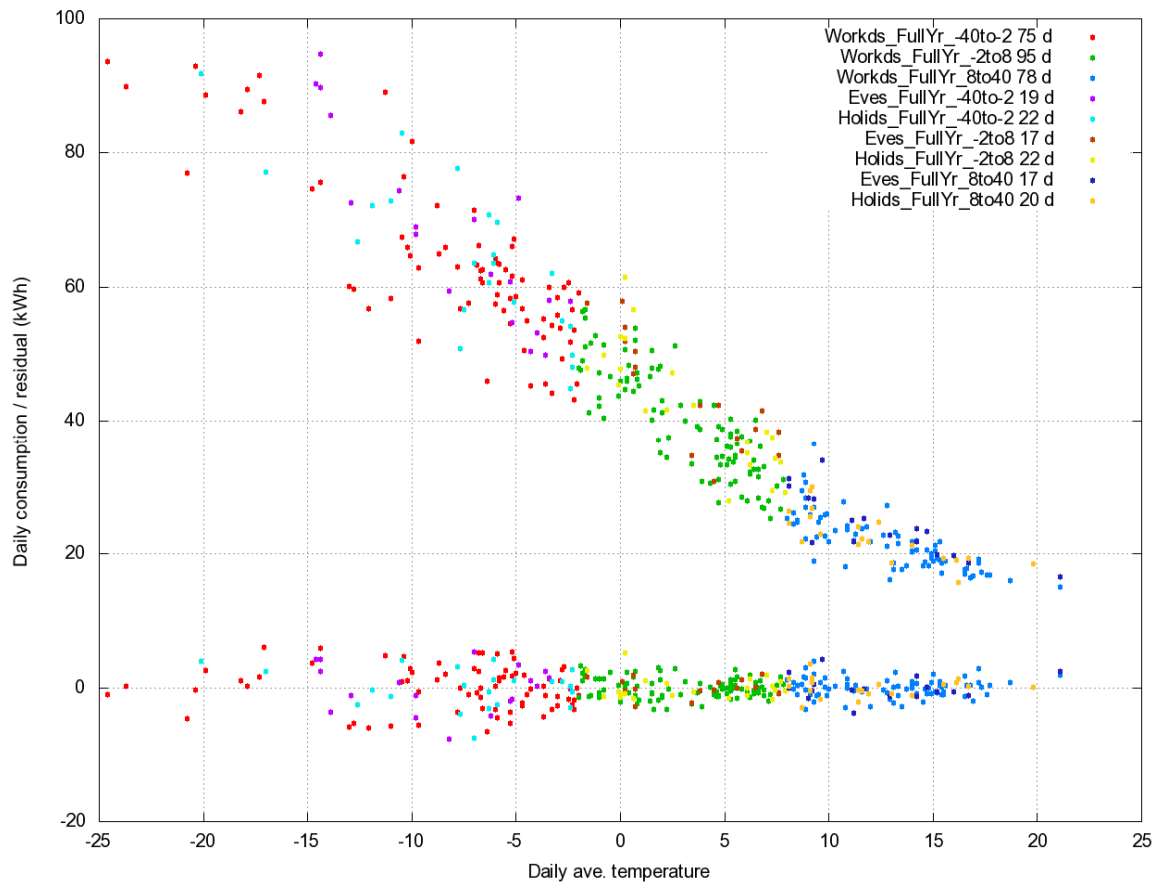
Kuva 30: Keskimääräisen varaavan sähkölämmittäjän tunneittainen mitattu (sinisellä) sekä ennustettu kulutus (punainen) ajalta 1-10.12.2010, kun päivät oli jaettu kolmeen lämpötila-alueeseen kohdista +8 ja -2 astetta ja kolmeen päivätyyppiin. Musta katkoviiva kuvaa edellisen 40 tunnin lämpötilakeskiarvoa, ja vihreällä katkoviivalla on kuvaan merkitty vuoden 2009 kulutus samalta ajalta.

#Date	h	real	predict	dif	dif %	stDev	pred+2*dev	T_short	T_long	Weekday
23/01/2010	23	9.07	8.25	-0.821	-9.1	0.358	8.96	-28.0	-27.7	Sat
28/01/2010	23	8.84	8.38	-0.461	-5.2	0.547	9.48	-22.7	-21.7	Thu
09/01/2010	23	8.81	8.10	-0.705	-8.0	0.358	8.82	-19.7	-22.3	Sat
07/01/2010	23	8.69	9.09	0.397	4.6	0.547	10.18	-29.7	-30.5	Thu
29/01/2010	23	8.63	7.93	-0.698	-8.1	0.547	9.02	-18.2	-19.9	Fri
21/01/2010	23	8.63	8.53	-0.092	-1.1	0.547	9.63	-24.2	-23.5	Thu
22/01/2010	23	8.59	8.69	0.105	1.2	0.547	9.79	-25.7	-26.0	Fri
05/01/2010	23	8.53	7.88	-0.654	-7.7	0.547	8.97	-17.7	-18.8	Tue
20/01/2010	23	8.47	7.95	-0.524	-6.2	0.547	9.04	-18.4	-15.7	Wed
27/01/2010	23	8.32	7.97	-0.351	-4.2	0.547	9.07	-18.6	-19.6	Wed
04/01/2010	23	8.29	8.08	-0.216	-2.6	0.547	9.17	-19.7	-19.3	Mon
02/01/2010	23	8.22	7.83	-0.383	-4.7	0.358	8.55	-21.7	-19.1	Sat
24/01/2010	0	8.15	9.59	1.442	17.7	0.500	10.59	-28.1	-27.8	Sun
29/01/2010	0	8.10	7.50	-0.598	-7.4	0.510	8.52	-22.5	-21.7	Fri
25/01/2010	23	8.10	7.83	-0.275	-3.4	0.547	8.92	-17.2	-17.5	Mon

Taulukko 9: Keskimääräisen varaavan sähkölämmittäjän vuoden 01/2010 25 kpl suurinta tuntikulutusta, kun päivät oli jaettu kolmeen lämpötila-alueeseen kohdista +8 ja -2 astetta ja kolmeen päivätyyppiin. Sarakkeessa real on toteutunut kulutus 2010, sarakeessa predict ennusteen antama kulutus ja sarakeessa pred+2\*dev 95 % luottamusarvioennuste. T\_short tarkoittaa edellisen 8 tunnin ja T\_long edellisen 40 tunnin lämpötilakeskiarvoa vuonna 2010.

## 7.4 Ilmalämpöpumppua käyttävä suora sähkölämmittäjä

Ilmalämpöpumppua käyttävien sähkölämmittäjien keskiarvolla oli 2008–2009 datan mukaan melko lineaarinen yhteys lämpötilaan. 2008–2009 talvi oli kuitenkin melko lämmin, joten ilmalämpöpumpun käyttäytyminen hyvin kylmässä lämpötilassa ei välttämättä tullut esille. Pelkona on, että hyvin kylminä päivinä lämpöpumpun hyötysuhde huononee ja näin tarvittava sähköteho saattaa kasvaa lineaarista ennustetta nopeammin.



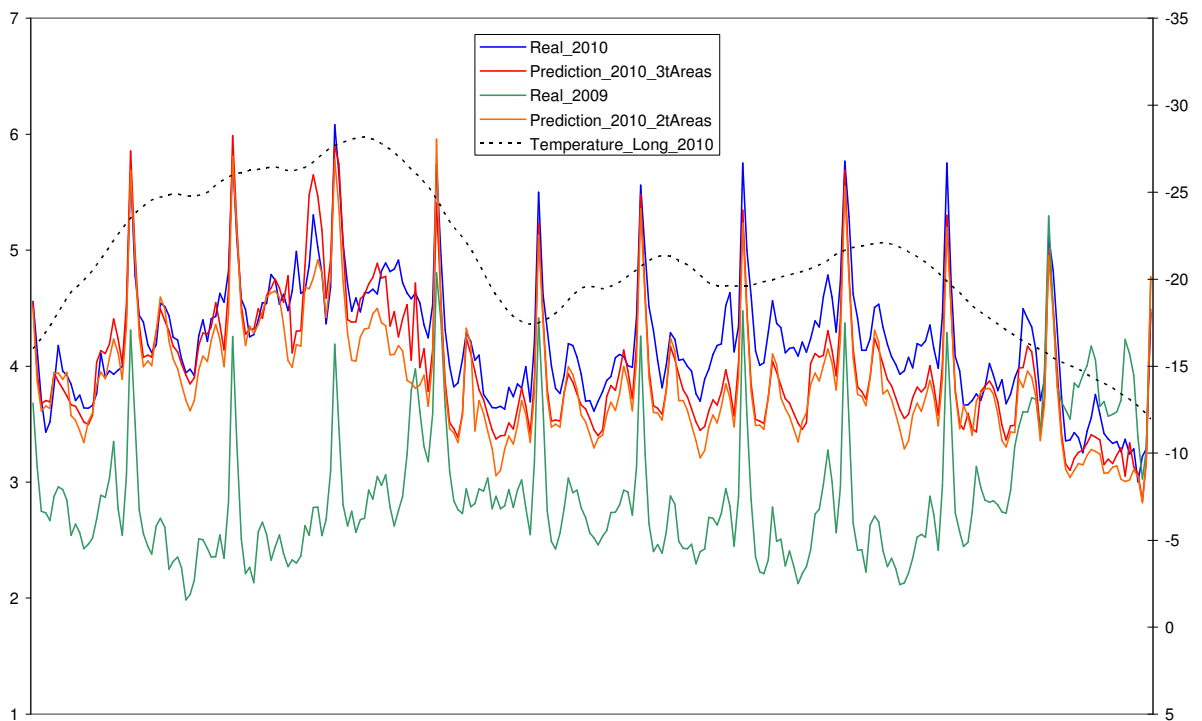
Kuva 31: Keskimääräisen ilmalämpöpumppua käyttävän suoran sähkölämmittäjän jako kolmeen lämpötila-alueeseen. Vuoden data 2008–2009. Y-akselille on merkitty päiväkulutus (kWh), ja x-akselin arvoiksi on valittu päivän lämpötilakeskiarvo, näkyy tässä ylempänä vinona joukkona pisteitä. Vuosi on jaettu kolmeen lämpötila-alueeseen kohdasta +8 ja -2 astetta, sekä kolmeen eri päivätyyppiin: esim. *Holds\_FullYear\_-40to-2* sisältää ne vuoden pyhäpäivät joiden lämpötila oli välillä -40...-2 astetta. Alempi joukko pisteitä esittää ennusteen tekemää virhettä eli residuaalia päiväkulutusten ennustamisessa eri lämpötiloissa.

Lämpöpumpun toimiessa huonosti asiakkaat saattavat myös kytkeä ylimääräisiä sähköpattereita verkkoon, jolloin lämmitykseen tarvittava sähköteho vielä enemmän. Sähköntarpeen ei-lineaarinen käyttäytyminen lämpötilan suhteen pitäisi näkyä päiväenergioissa, toisin kuin



varaavan sähkölämmittäjän tapauksessa jossa teho siirtyy tunnilta toiselle päiväenergian säilyessä samana. Kuvassa 31 ei kuitenkaan huomata ainakaan merkittävää epälineaarisuutta 2008–2009 datan perusteella. Kylminä päivinä hajonta on kuitenkin melko suurta, joten kylmien päivien käyttäytymistä on vaikea arvioida.

2010 tammikuun ennustetta ja mitattua kulutusta arvioitaessa kolmen lämpötila-alueen ennuste kuitenkin antoi paremman arvion kuin kahteen lämpötila-alueeseen perustuva ennuste. Ero ei ollut kovin suuri, mutta kahden lämpötila-alueen ennuste antoi tasaisesti liian matalia kulutusarvoja. 01/2010 kulutus oli ilmalämpöpumppua ja suoraa sähkölämmitystä käyttävillä asiakkailla keskimäärin 19,5 % korkeampi kuin 01/2009 kulutus. Tämä on suurempi arvo kuin aiemmin käsitellyillä sähkölämmitystä käyttävillä asiakkailla, ja sen perusteella voisi epäillä ei-lineaarista yhteyttä lämpötilan ja sähkönkulutuksen välillä. Ilmalämpöpumppua käyttävien asiakkaiden lämmitettävät pinta-alat ovat kuitenkin keskimäärin huomattavasti suurempia kuin muiden tässä käsiteltyjen sähkölämmittäjien. Tämä näkyy suurina päiväkulutuksina, kuten nähdään vertaamalla kuvaa 31 kuvaan 5 ja liitteeseen 1. Ilmalämpöpumppua käyttäviä asiakkaita ei voida suoraan verrata muihin tässä tutkittuihin sähkölämmittäjiin.



*Kuva 32: Keskimääräisen ilmalämpöpumppua käyttävän suoran sähkölämmittäjän tunneittainen mitattu sekä ennustettu kulutus ajalta 21-31.12.2010. Punaisella kolmen ja oranssilla kahden lämpötila-alueen ennuste. Musta katkoviiva kuvaa edellisen 40 tunnin lämpötilakeskiarvoa, ja vihreällä katkoviivalla on kuvaan merkitty vuoden 2009 kulutus samalta ajalta.*

Kuvassa 32 nähdään esimerkki ilmalämpöpumppua käyttävän sähkölämmittäjän ennusteesta kolmen ja kahden lämpötila-alueen avulla tehtynä. Jakamalla vuoden päivät kolmeen lämpötila-alueeseen pystytään paremmin hallitsemaan epälineaarinen riippuvuus lämpötilan ja sähkökulutuksen välillä, kuten huomattiin aikaisemmin varaavan sähkölämmittäjän tapauksessa. Ilmalämpöpumpun tapauksessa epälineaarisuus on paljon heikompaa, mutta ennusteen kannalta pahempaan suuntaan eli maksimikulutusennusteesta saattaa tulla liian alhainen. On kuitenkin vaikeaa sanoa, riippuuko epälineaarisuus itse asiassa lämpöpumpusta vai siitä että niitä käyttävillä asiakkailla on keskimäärin suuri lämmitettävä pinta-ala. Kolmeen lämpötila-alueeseen jakamalla pystytään joka tapauksessa paremmin ennustamaan yllättävät kylmien päivien kulutukset. Liitteessä 15 on esitettyä kuvan 32 tapaan 1-20.1.2010 eri tavoilla tehty ennuste ja mitattu kulutus.

## 8 Yhteenveto ja jatkotutkimusehdotukset

Kotitalouksien tunneittaisen sähkönkulutuksen ennustaminen onnistui melko hyvin. Pienien asiakasmäärien summakulutuksen ennustaminen on hankalaa, mutta ainakin periaatteessa tässä työssä esitelty ohjelma pystyy ennustamaan yleisimpien lämmitystavan mukaan eroteltujen asiakasryhmien sähkönkulutuksen melko tarkasti. Myös erityyppisistä asiakkaista koostuvien pienten asiakasryhmien summakulutus pystytään ennustamaan, mutta tässä työssä ei vielä pystytty antamaan hyvin perusteltua arviota tarvittavasta asiakkaiden minimimäärästä. Erikokoisten pienten asiakasryhmien summakulutuksen analysointi onkin tärkeimpiä jatkotutkimuksen aiheita tämän työn jälkeen.

Toinen auki jäänyt kysymys on valinta vuoden jakamisesta kahteen tai kolmeen lämpötila-alueeseen. Jako arki-, aatto- ja pyhäpäiviin näytti olevan ainut järkevä jako, jos halutaan kaikille asiakastyypeille sopiva mahdollisimman automaattinen ennusteohjelma. Jakamalla vuosi kahteen lämpötila-alueeseen saatiin yleensä melko hyvä ennuste, mutta lämpötilan mukaan ei-lineaarisesti käyttäytyvien asiakastyypin ennustamisessa jako kolmeen alueeseen antoi paremman tuloksen. Varsinkin varaavaa sähkölämmitystä käyttävien asiakkaiden kello 1 ja 2 ei-lineaarinen lämpötilariippuvuus pystytään mallintamaan paljon paremmin kolmea lämpötila-aluetta käyttämällä.

Kahteen lämpötila-alueeseen jakaminen varmistaa, että kuhunkin lineaarisesti käsiteltävään päiväryhmään jää regressioanalyysiä varten riittävä määrä päiviä. Tässä työssä on käytetty yhden vuoden mittausdataa ennusteen muodostamiseen, mutta on todennäköistä että tuntimitattua dataa tullaan säilyttämään pitemmältä ajalta. Tällöin myös kolmeen lämpötila-alueeseen jakamalla saadaan helpommin riittävä määrä mittausdataa sähkönkulutuksen lämpötila- ja päivänpituuskorrelaation löytämiseksi, vaikka yhden vuoden yhteen päiväryhmään voikin jäädä melko vähän päiviä.

Tässä työssä käsitelty ennustevirheen hajonta kuvaa tietyn päiväryhmän tietyn kellonajan eri päivien hajontaa jota ei pystytty selittämään lämpötilan tai päivänpituuden avulla. Olettamalla normaalijakauma ja riippumattomuus selittävistä muuttujista, voidaan laskea esimerkiksi 95 % luottamustaso. Tällöin voidaan antaa huipputehoarvio tietylle tunnille niin, että 97,5 % todennäköisyydellä sähkönkulutus ei ylitä tätä tasoa. Tällöin kuitenkin oletetaan että asiakkaan sähkönkäyttö säilyy tulevaisuudessa samanlaisena, ja ainoastaan lämpötila muuttuu.

Edellä esitetyn lisäksi oletetaan, että sekä lämpötilan että päivänpituuden vaikutus sähkönkulutukseen on onnistuttu määrittämään oikein. Jos määritettyyn päiväryhmään kuuluu hyvin vähän tunteja ja riippuvuus esimerkiksi lämpötilan ja sähkönkulutuksen välillä ei ole selkeä, jää epävarmaksi kuvaako matemaattisen analyysin antama lineaarinen suora lämpötilan vaikutuksen oikein. Tällaisessa tilanteessa sovituksen hyvyttä kuvaava  $R^2$  on todennäköisesti pieni, mutta vielä paremmin tilannetta kuvaaisi sovitetun suoran jyrkkyyden luottamustaso. Tämä on esitetty viitteessä [9], kohdassa Confidence or prediction interval of a regression line. Tätä soveltamalla pystyttäisiin arvioimaan lämpötilariippuvuuden luottamustaso, jolloin voidaan yrittää estää väärin analysoidun lämpötilariippuvuuden aiheuttamia karkeita virheitä kun mallia käytetään ennustamaan kylmempiä päiviä kuin mitä alkuperäisessä datassa on ollut mukana.

Jos käytössä on tuntimitattua dataa usealta vuodelta, voidaan yrittää analysoida vuosittainen kasvutrendi. Trendi voisi olla lämpötilan ja päivänpituuden ohella kolmas selittävä tekijä. Tätä voitaisiin testata yksinkertaisimmillaan olettamalla lineaarinen kasvu vuodesta toiseen ja järjestää vuodet aikajärjestykseen esimerkiksi yhdestä kuuteen, jos käytössä kuuden vuoden data. Näin meillä olisi yksi uusi selittävä tekijä. On kuitenkin epävarmaa kuinka samanlaisina pienten asiakasryhmien kulutus säilyy usean vuoden aikana, mutta trendin mukaan ottoa malliin voitaisiin kuitenkin testata. Näin voitaisiin määrittää tiettyjen alueiden tai tiettyjen asiakastyypien vuotuinen sähkönkulutuksen kasvuprosentti.

Tehty ohjelma on suunniteltu käsittelemään asiakasryhmän summaa. Tällöin asiakkaiden välinen hajonta näkyy summan hajonnassa, joten yksittäisten asiakkaiden hajontaa ei tarvitse laskea. On kuitenkin mahdollista että jossain tilanteissa haluttaisiin esimerkiksi summata yhteen jo käsiteltyjen jakelumuuntajien kulutuksia. Kappaleessa 4.4 on käsitelty summan hajonnan muodostumista yksittäisistä asiakkaista tai aliryhmistä, mutta sitä ei ole implementoitu ohjelmaan. Esitettyjen periaatteiden mukaisesti voidaan kuitenkin haluttaessa yrittää tarkemmin käsitellä hajonnan muodostumista. Toisaalta useita muuntajia summattaessa voidaan myös koota kakkien käsiteltävien muuntajien takana olevat asiakkaat yhdeksi ryhmäksi ja käsitellä näiden summakulutus yhtenä aikasarjana tässä työssä esitettyjen periaatteiden mukaisesti; tällöin pärjätään tässä työssä esitellyllä ohjelmalla.

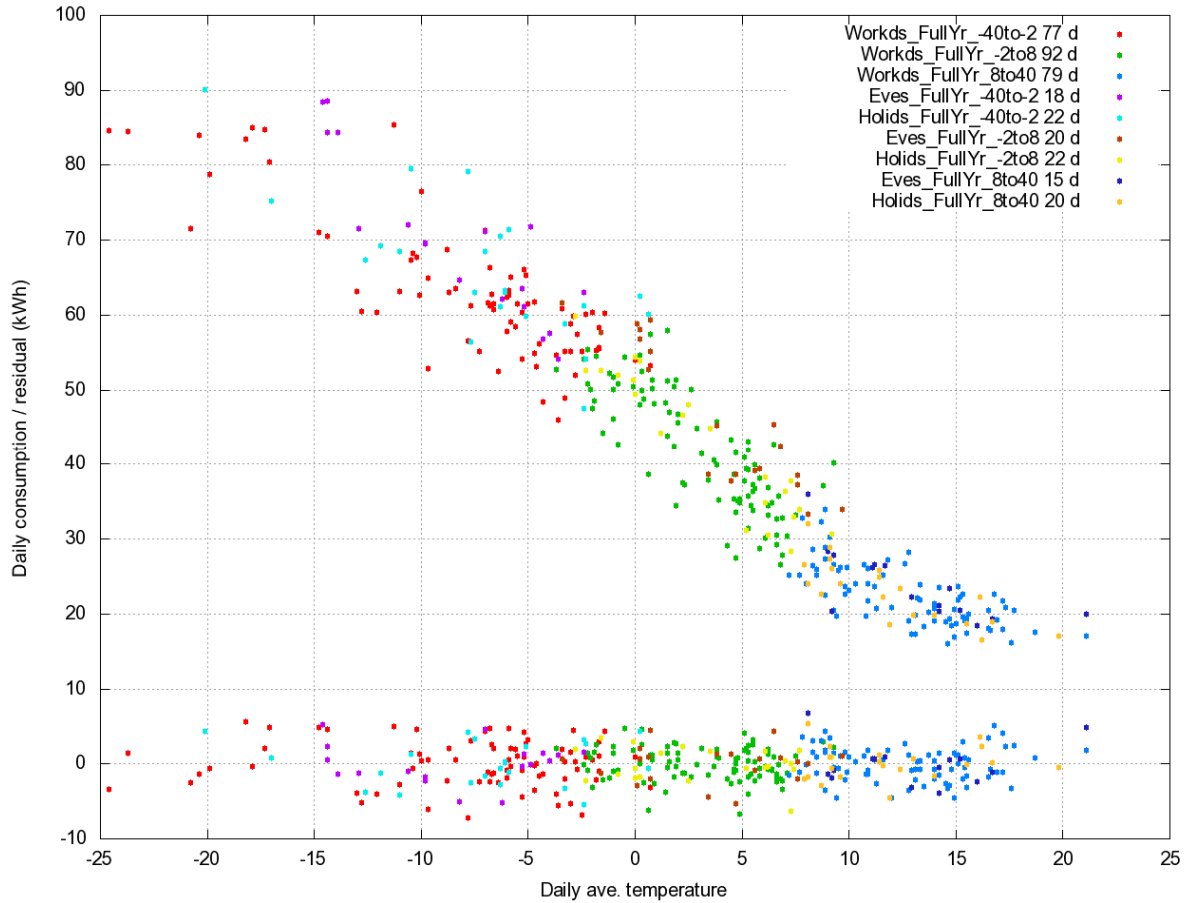
Edellä esitetyt jatkotutkimusaiheet koskevat ennustamisen teoreettista puolta, sekä ohjelman laskentaa varten tehtäviä määrittelyitä ja yksinkertaistuksia: esimerkiksi vuoden jako päiväryhmiin. Tämän lisäksi, ohjelman käyttöönottamiseksi esimerkiksi verkostolaskennan apuna, tulisi parantaa ohjelman nopeutta ja varmistaa toimivuus eri tilanteissa testien avulla.

## Lähdeluettelo

- [1] A model comparison for daylength as function of latitude and day of year, William C. Forsythe, Edward J. Rykiel Jr., Randal S. Stahl, Hsin-i Wu, Robert M. Schoolfield, Ecological Modelling 80 (1995) 87-95.
- [2] <http://www.heavens-above.com/selecttown.asp?CountryID=FI> (etsitty Kainuu).
- [3] Clemson University physics opetusmateriaali lineaarisesta regressiosta: <http://phoenix.phys.clemson.edu/tutorials/regression/index.html> (viittaus tarkistettu 1.4.2010).
- [4] Helsingin yliopiston kasvatustieteen laitoksen tilastomenetelmien jatkokurssin opetusmoniste, Karma, Kai & Komulainen, Erkki. Toinen laitos (versio 2.2, 1.1.2002). Helsingin yliopisto, kasvatustieteen laitos.
- [5] Lämpötilan vaikutus sähkön kulutukseen, Matti Koivisto, Erikoistyö, Sähkötekniikan laitos, Aalto-yliopiston teknillinen korkeakoulu.
- [6] <http://en.wikipedia.org/wiki/Multicollinearity> (viittaus tarkistettu 1.4.2010).
- [7] [http://en.wikipedia.org/wiki/68-95-99.7\\_rule](http://en.wikipedia.org/wiki/68-95-99.7_rule) (viittaus tarkistettu 1.4.2010).
- [8] [http://en.wikipedia.org/wiki/Standard\\_deviation](http://en.wikipedia.org/wiki/Standard_deviation) (viittaus tarkistettu 1.4.2010).
- [9] [http://en.wikipedia.org/wiki/Regression\\_analysis](http://en.wikipedia.org/wiki/Regression_analysis) (viittaus tarkistettu 1.4.2010).
- [10] Sähkönjakelutekniikka, Erkki Lakervi & Jarmo Partanen, Otatieto/Gaudeamus 2008.
- [11] Construction of SARIMAX models using MATLAB, Mat-2.4108 Independent research projects in applied mathematics, Antti Savelainen, 25.9.2009.
- [12] Sähkönkulutuksen mittauksen uudistus, elinkeinoministeri Mauri Pekkarisen tiedotustilaisuus 5.2.2009, Työ- ja elinkeinoministeriö.
- [13] A method for assessing the maximum load flows in distribution systems, Mati Meldorf and Toomas Täht from Tallinn University of Technology and Matti Lehtonen, Anssi Seppälä and Mikko Jalonen from VTT Energy, Energy Systems, PSCC 1999.
- [14] <http://en.wikipedia.org/wiki/Variance> (viittaus tarkistettu 1.4.2010).
- [15] <http://en.wikipedia.org/wiki/Covariance> (viittaus tarkistettu 1.4.2010).
- [16] Zeller's congruence Wikipediassa: [http://en.wikipedia.org/wiki/Zeller%27s\\_congruence](http://en.wikipedia.org/wiki/Zeller%27s_congruence) (viittaus tarkistettu 1.4.2010).
- [17] Suomen juhlapäivät Wikipediassa: [http://fi.wikipedia.org/wiki/Suomen\\_juhlap%C3%A4iv%C3%A4t](http://fi.wikipedia.org/wiki/Suomen_juhlap%C3%A4iv%C3%A4t) (viittaus tarkistettu 1.4.2010).

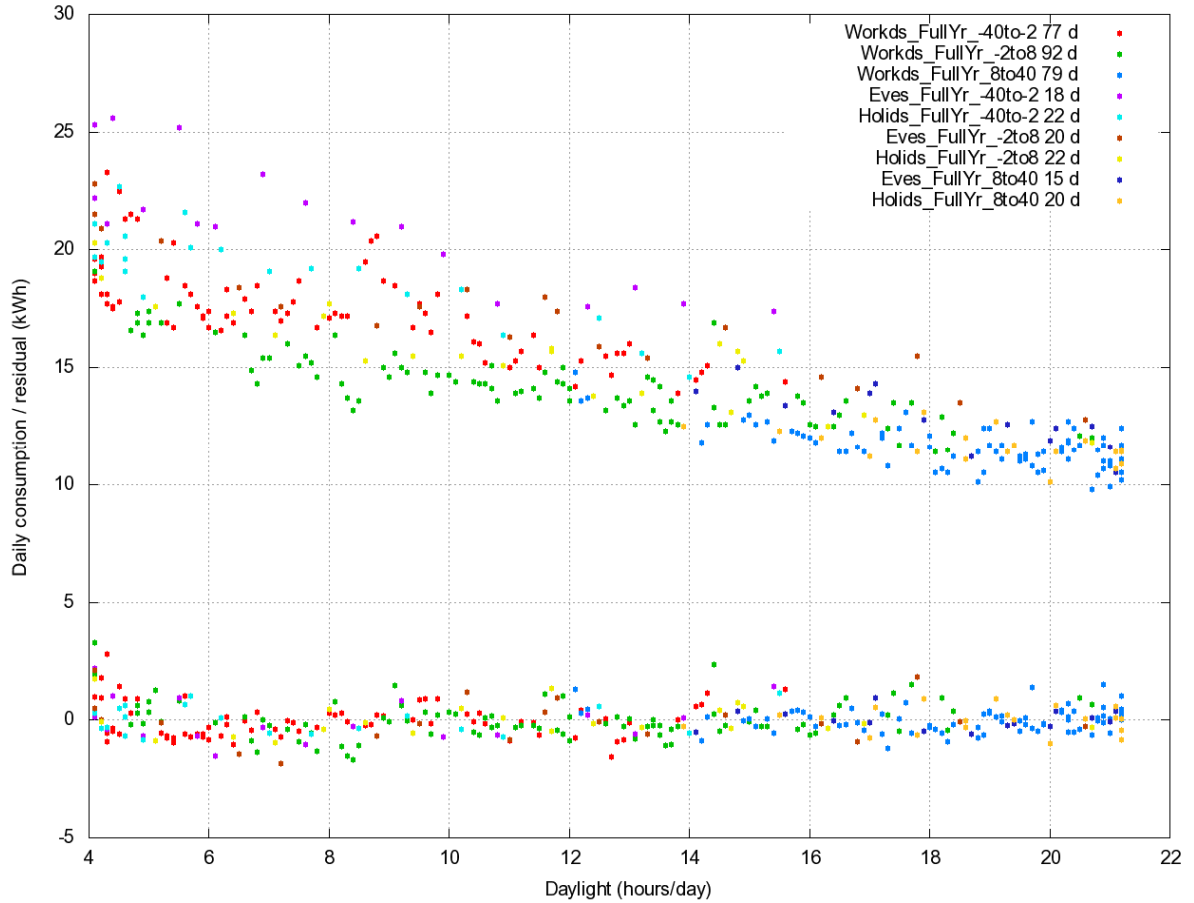
- 
- [18] Pääsiäisen päivämäärä eri vuosina Suomessa käytettävässä kalenterissa Wikipediassa:  
<http://fi.wikipedia.org/wiki/P%C3%A4%C3%A4si%C3%A4inen> (viittaus tarkistettu 1.4.2010).
- [19] [http://www.curvefit.com/linear\\_regression.htm](http://www.curvefit.com/linear_regression.htm) (viitattu 14.4.2010)

## Liite 1: Varaavan sähkölämmittäjän jako kolmeen lämpötila-alueeseen



*Keskimääräisen varaavan sähkölämmittäjän jako kolmeen lämpötila-alueeseen. Y-akselille on merkitty päiväkulutus (kWh), ja x-akselin arvoiksi on valittu päivän lämpötilakeskiarvo, näkyy tässä ylempänä vinona joukkona pisteitä. Vuosi on jaettu kolmeen lämpötila-alueeseen kohdasta +8 ja -2 astetta, sekä kolmeen eri päivätyyppiin: esim. `Holids_FullYear_-40to-2` sisältää ne vuoden pyhäpäivät joiden lämpötila oli välillä -40..-2 astetta. Alempi joukko pisteitä esittää ennusteen tekemää virhettä eli residuaalia päiväkulutusten ennustamisessa eri lämpötiloissa.*

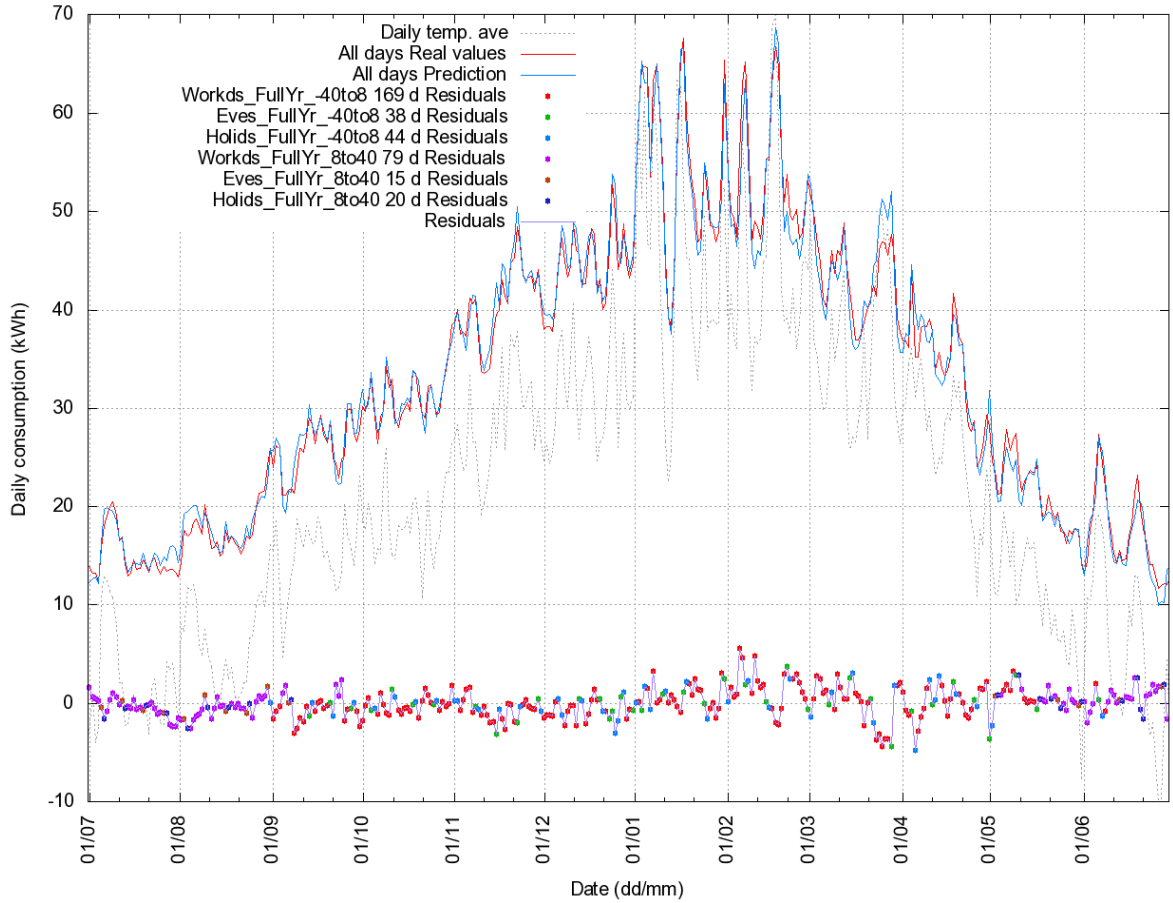
## Liite 2: Päivänpituuden vaikutus öljylämmittäjän kulutukseen



Keskimääräisen öljylämmitystä käyttävän asiakkaan päiväkulutus (kWh), kun x-akselin arvoiksi on valittu päivän pituus tunteina. Kulutus näkyy tässä ylempänä vinona joukkona pisteitä. Vuosi on jaettu kolmeen lämpötila-alueeseen kohdista +8 ja -2 astetta, sekä kolmeen eri päivätyyppiin: esim. *Holids\_FullYear\_-40to-2* sisältää ne vuoden pyhäpäivät joiden lämpötila oli välillä -40...-2 astetta. Alempi joukko pisteitä kuvaa ennusteen tekemää virhettä eli residuaalia päiväkulutusten ennustamisessa eri päivän pituuden arvoilla.

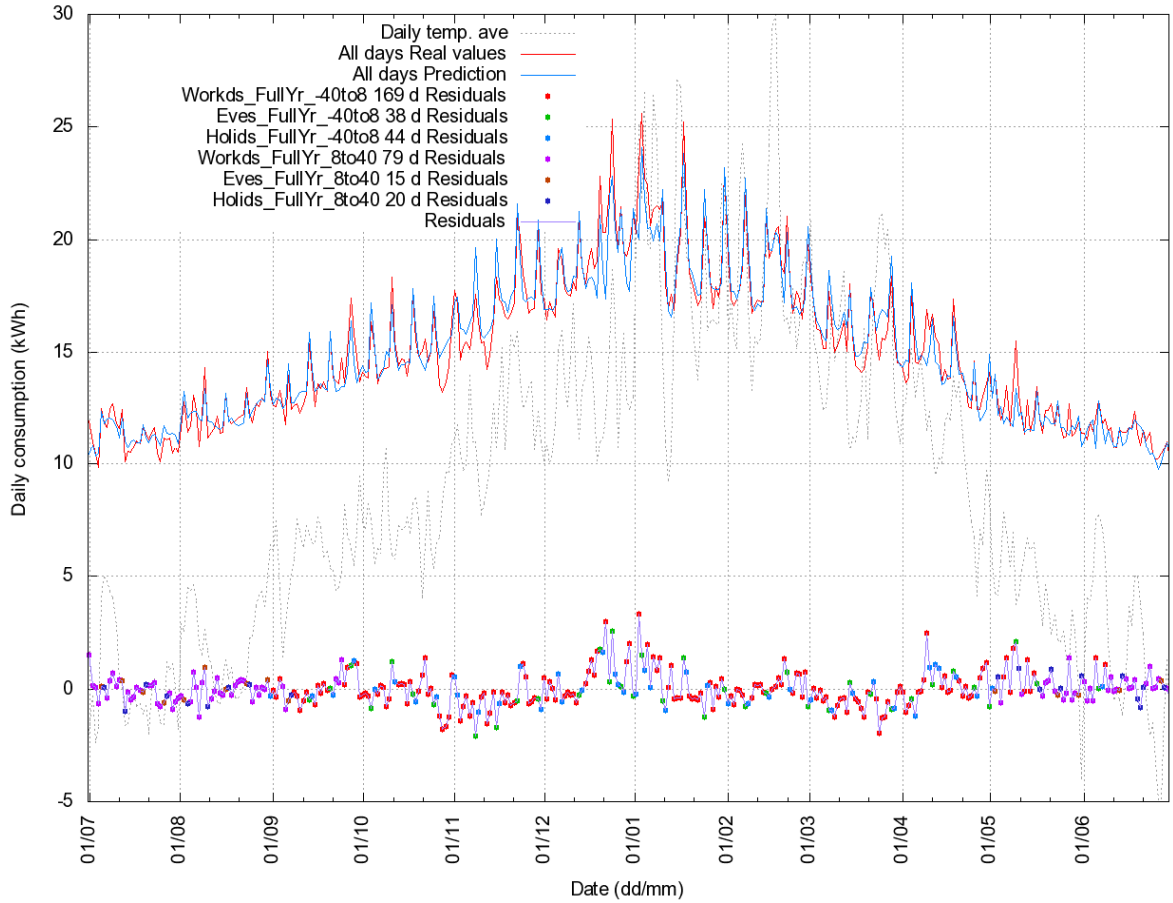


### Liite 3: Suoran sähkölämmittäjän mitattu ja ennustettu kulutus



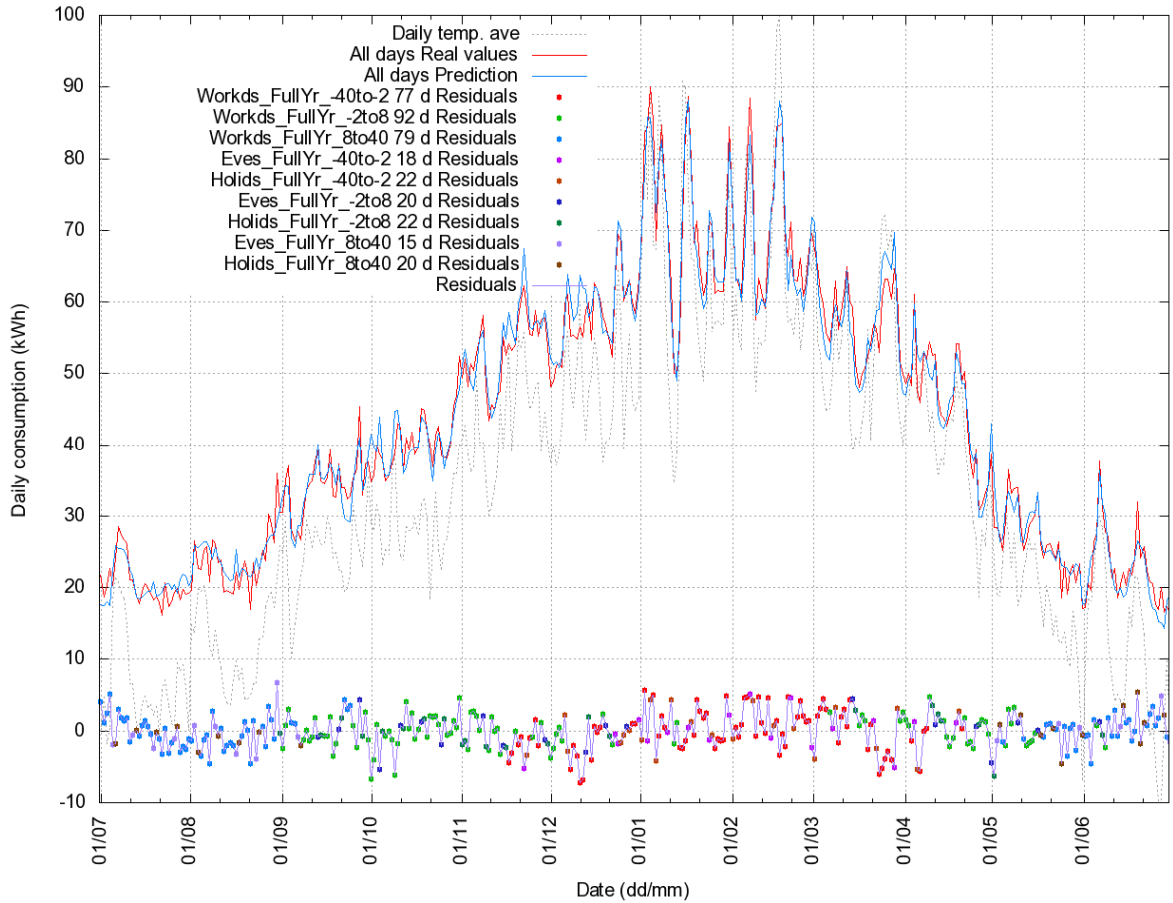
*Keskimääräisen suoran sähkölämmittäjän mitattu (punainen) ja ennustettu (sininen) päiväkulutus vuoden aikana, kun vuosi on jaettu kahteen lämpötila-alueeseen kohdasta +8 astetta, sekä kolmeen eri päivätyyppiin. Kuvan alaosassa näemme ennusteen tekemän virheen eli residuaalin vuoden eri päivinä.*

## Liite 4: Öljylämmittäjän mitattu ja ennustettu kulutus



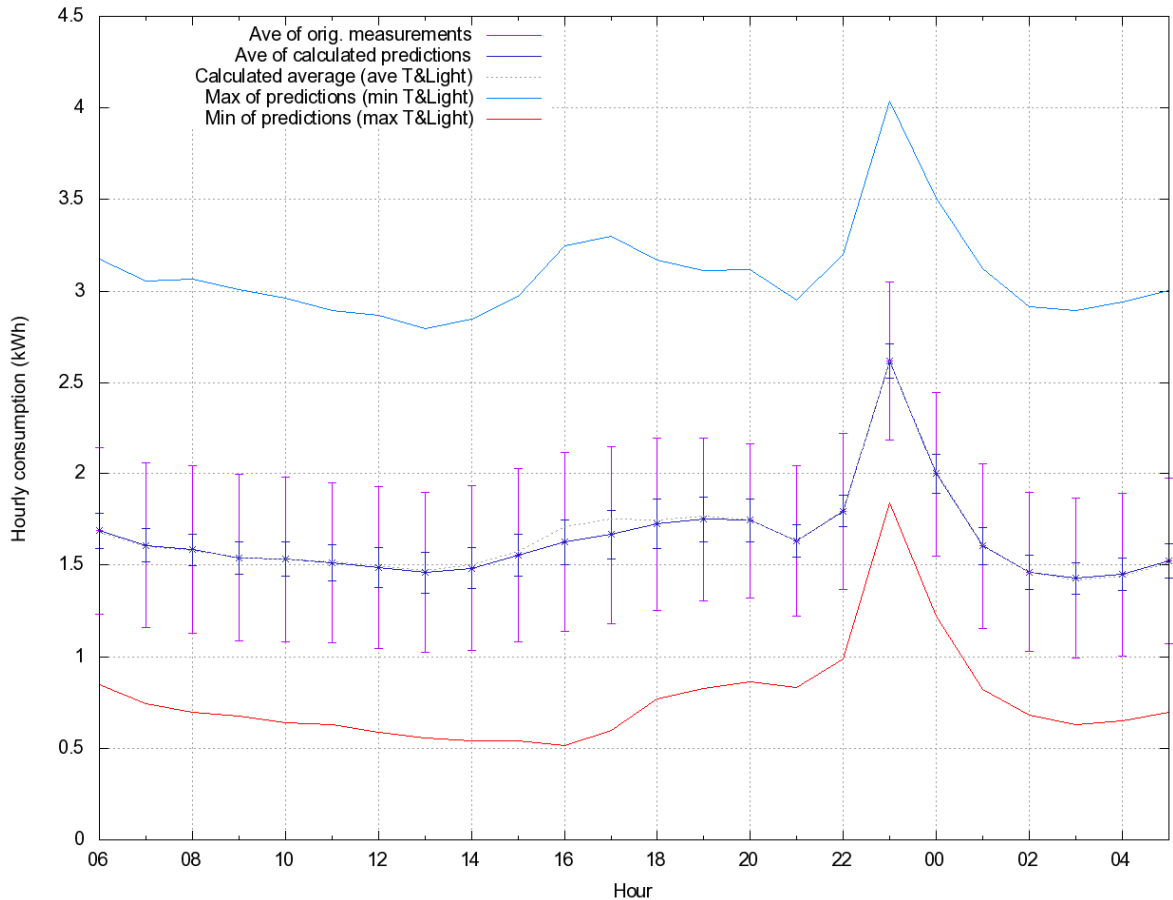
*Keskimääräisen öljylämmittäjän mitattu (punainen) ja ennustettu (sininen) päiväkulutus vuoden aikana, kun vuosi on jaettu kahteen lämpötila-alueeseen kohdasta +8 astetta, sekä kolmeen eri päivätyyppiin. Kuvan alaosassa näemme ennusteen tekemän virheen eli residuaalin vuoden eri päivinä.*

## Liite 5: Varaavan sähkölämmittäjän mitattu ja ennustettu kulutus kun jako kolmeen lämpötila-alueeseen



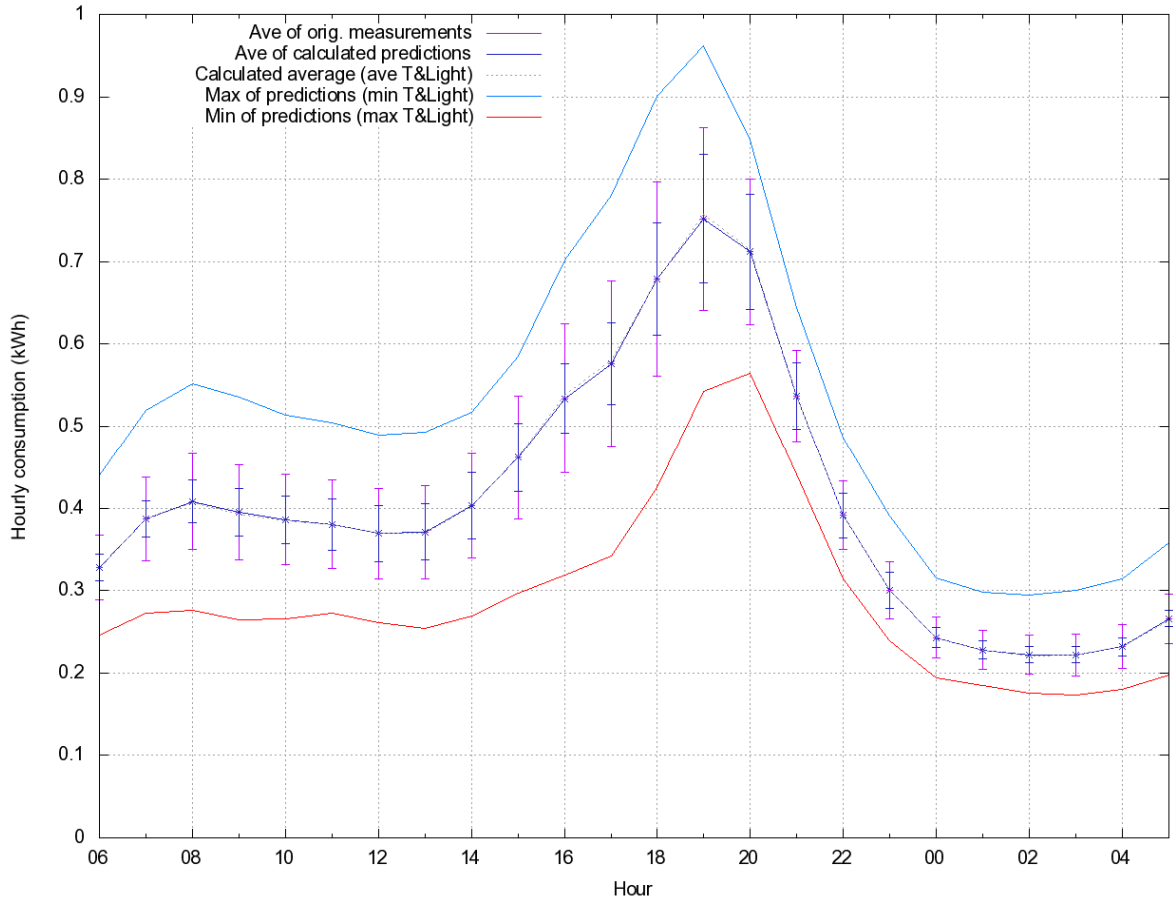
Keskimääräisen varaavan sähkölämmittäjän mitatun (punainen) ja ennustetun (sininen) päiväkulutuksen vuoden aikana, kun vuosi on jaettu kolmeen lämpötila-alueeseen kohdasta +8 ja -2 astetta, sekä kolmeen eri päivätyyppiin. Kuvan alaosassa näemme ennusteen tekemän virheen eli residuaalin vuoden eri päivinä.

## Liite 6: Suoran sähkölämmittäjän kylmien arkipäivien rakenne



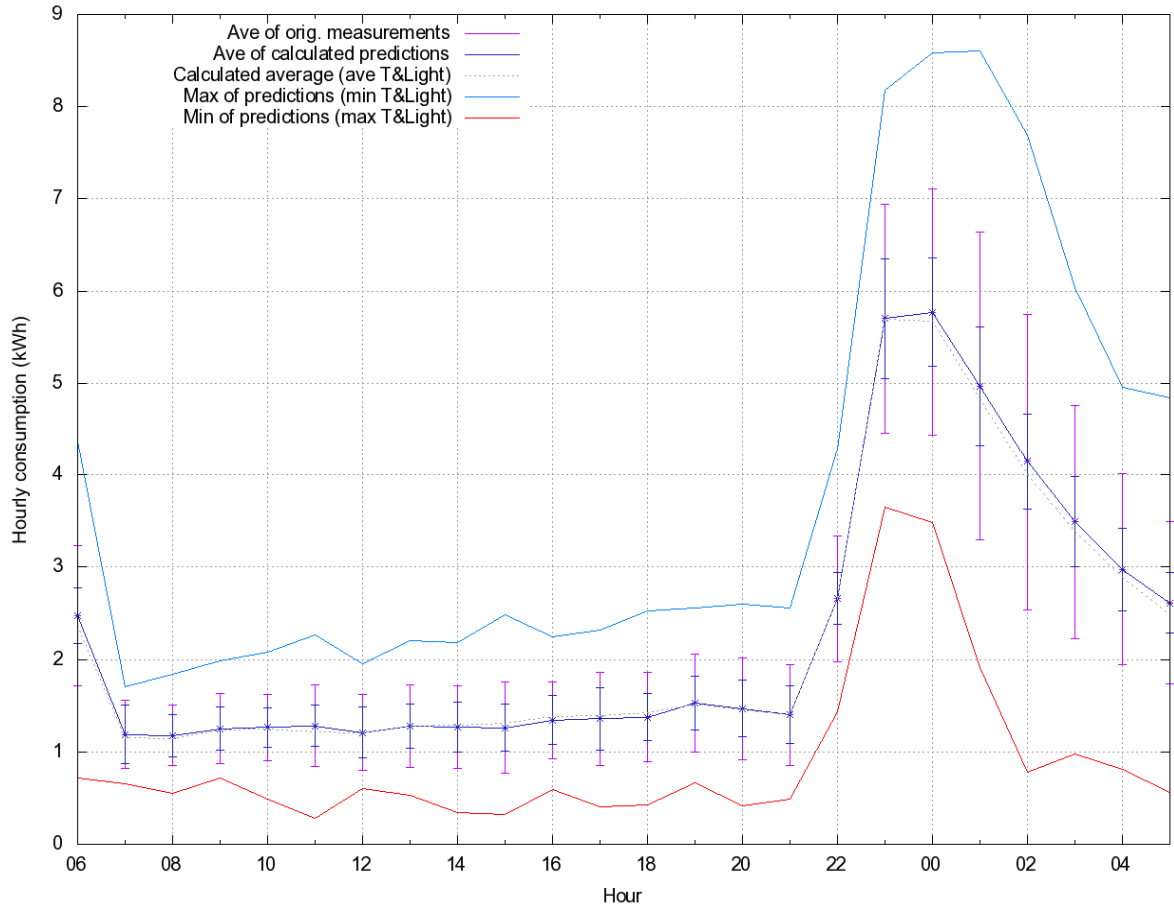
Tummansinisellä keskimääräisen suoran sähkölämmittäjän päivärakenteen arkipäivänä, kun päivän keskiarvolämpötila on alle +8 astetta. Kuvaan on merkitty myös hajonta eri tunneilla ennen sovitusta tummansinisellä ja lilalla sovituksen jälkeen: pieni hajonta kuvaa hyvin onnistunutta sovitusta. Vaaleansinisellä on kuvattu mallin antama ennuste jos samaan aikaan olisi sekä erittäin kylmää että pimeää (käsitellyn aikavälin toteutuneet minimi), eli se kuvaa mallin antamaa suurinta mahdollista päiväkulutusta. Punainen käyrä kuvaa samalla periaatteella mallin antaman minimipäiväkulutuksen.

## Liite 7: Kaukolämpöasiakkaan kylmien arkipäivien rakenne



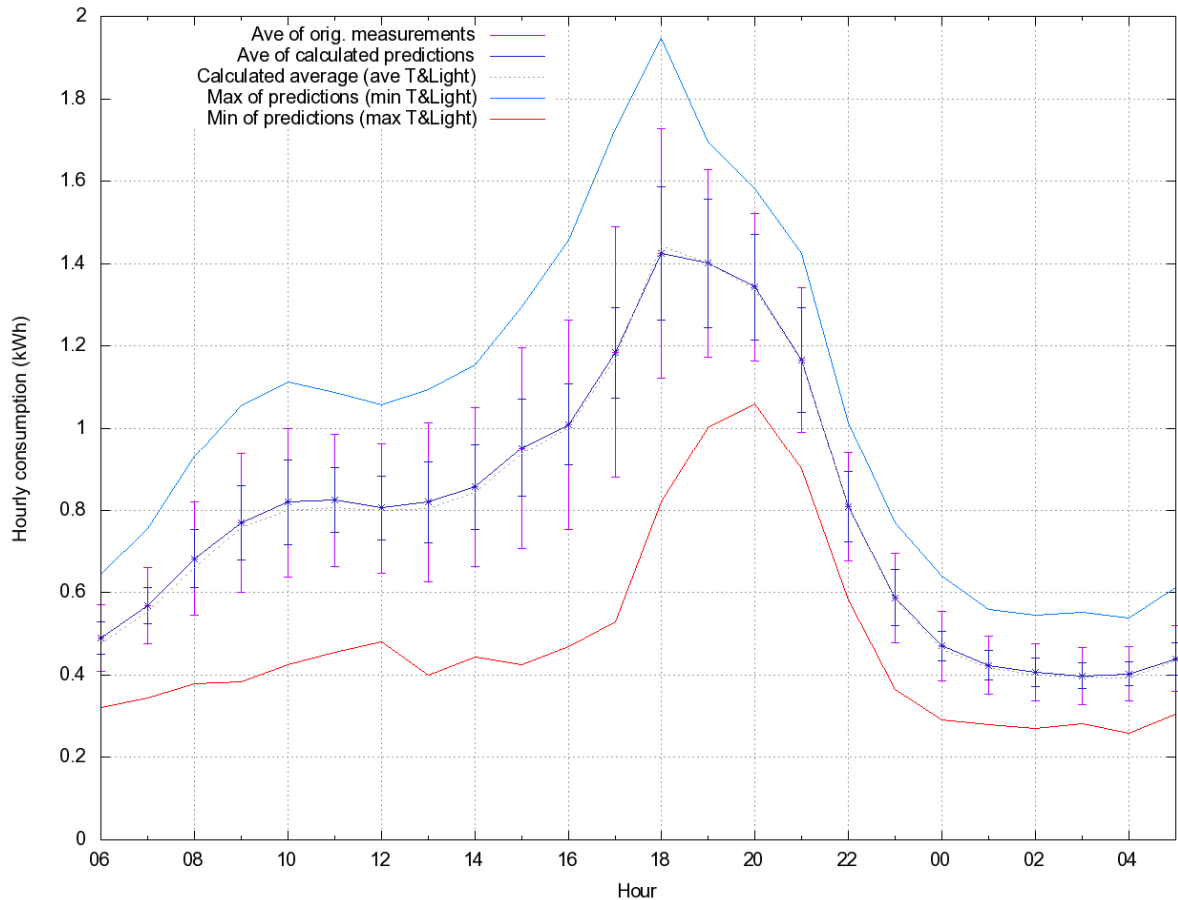
*Tummansinisellä keskimääräisen kaukolämpöä käyttävän asiakkaan päivärakenteen arkipäivänä, kun päivän keskiarvolämpötila on alle +8 astetta. Kuvaan on merkitty myös hajonta eri tunneilla ennen sovitusta tummansinisellä ja lilalla sovituksen jälkeen: pieni hajonta kuvaa hyvin onnistunutta sovitusta. Vaaleansinisellä on kuvattu mallin antama ennuste jos samaan aikaan olisi sekä erittäin kylmää että pimeää (käsitellyn aikavälin toteutuneet minimi), eli se kuvaa mallin antamaa suurinta mahdollista päiväkulutusta. Punainen käyrä kuvaa samalla periaatteella mallin antaman minimipäiväkulutuksen.*

## Liite 8: Varaavan sähkölämmittäjän kylmien pyhäpäivien rakenne



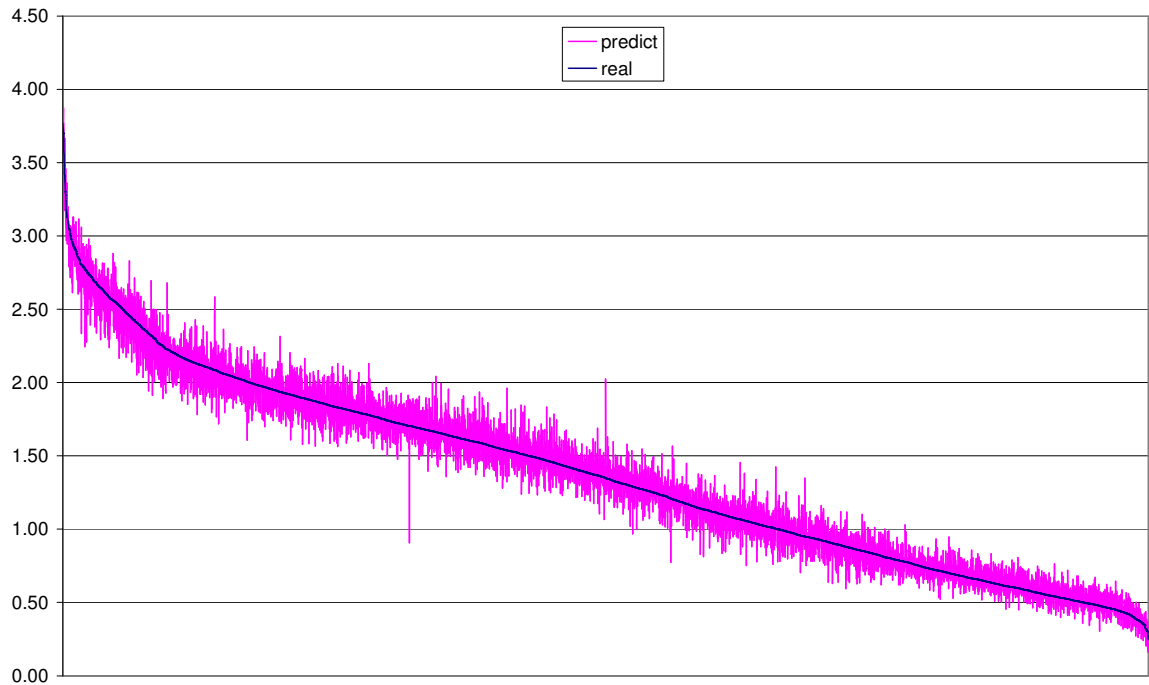
*Tummansinisellä keskimääräisen varaavan sähkölämmittäjän päivärakenteen pyhäpäivänä, kun päivän keskiarvolämpötila on alle +8 astetta. Kuvaan on merkitty myös hajonta eri tunneilla ennen sovitusta tummansinisellä ja lilalla sovituksen jälkeen: pieni hajonta kuvaa hyvin onnistunutta sovitusta. Vaaleansinisellä on kuvattu mallin antama ennuste jos samaan aikaan olisi sekä erittäin kylmää että pimeää (käsitellyn aikavälin toteutuneet minimi), eli se kuvaa mallin antamaa suurinta mahdollista päiväkulutusta. Punainen käyrä kuvaa samalla periaatteella mallin antaman minimipäiväkulutuksen.*

## Liite 9: Öljylämmittäjän kylmien aattopäivien rakenne



*Tummansinisellä keskimääräisen öljylämmittäjän päivärakenteen aattopäivänä, kun päivän keskiarvolämpötila on alle +8 astetta. Kuvaan on merkitty myös hajonta eri tunneilla ennen sovitusta tummansinisellä ja lilalla sovituksen jälkeen: pieni hajonta kuvaa hyvin onnistunutta sovitusta. Vaaleansinisellä on kuvattu mallin antama ennuste jos samaan aikaan olisi sekä erittäin kylmää että pimeää (käsitellyn aikavälin toteutuneet minimi), eli se kuvaa mallin antamaa suurinta mahdollista päiväkulutusta. Punainen käyrä kuvaa samalla periaatteella mallin antaman minimipäiväkulutuksen.*

## Liite 10: Suoran sähkölämmittäjän kulutushuiput 2008-2009



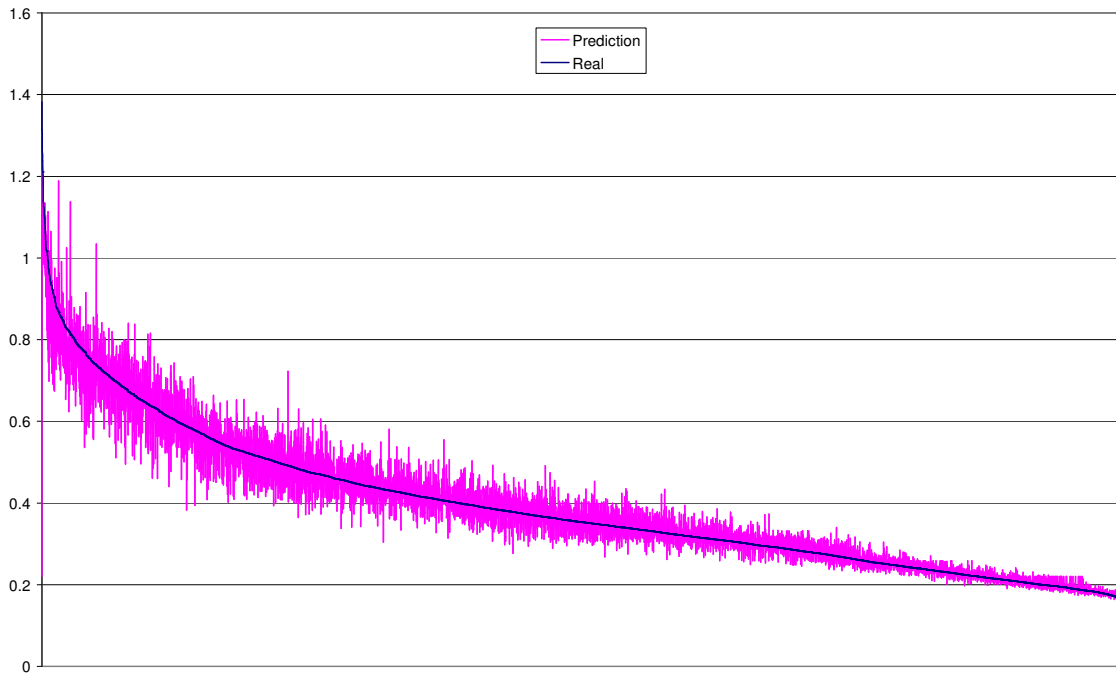
Keskimääräisen suoran sähkölämmittäjän tuntikulutukset järjestettynä alkuperäisen mittauksen suuruuden mukaan sinisellä sekä liialla ennusteen antamat arvot samoille tunneille. Käytössä oli perusjako kuuteen päiväryhmään.

#Date	h	real	predict	dif	dif_ %	stDev	pred+2*dev	T_short	T_long	Weekday
16/01/2009	23	3.77	3.78	0.012	0.3	0.094	3.97	-21.3	-21.3	Fri
07/01/2009	23	3.73	3.62	-0.109	-2.9	0.094	3.81	-20.4	-16.0	Wed
17/02/2009	23	3.73	3.87	0.148	4.0	0.094	4.06	-25.7	-24.0	Tue
04/01/2009	23	3.70	3.67	-0.034	-0.9	0.143	3.95	-19.2	-16.2	Sun
16/02/2009	23	3.70	3.77	0.071	1.9	0.094	3.96	-24.4	-20.9	Mon
08/01/2009	23	3.70	3.62	-0.079	-2.1	0.094	3.81	-18.2	-18.0	Thu
17/01/2009	23	3.65	3.48	-0.166	-4.6	0.106	3.69	-15.2	-17.5	Sat
06/02/2009	23	3.57	3.49	-0.081	-2.3	0.094	3.68	-17.0	-15.9	Fri
15/01/2009	23	3.55	3.49	-0.060	-1.7	0.094	3.68	-18.6	-12.5	Thu
30/01/2009	23	3.50	3.22	-0.281	-8.0	0.094	3.41	-9.6	-8.2	Fri
05/02/2009	23	3.50	3.18	-0.327	-9.3	0.094	3.36	-12.0	-8.4	Thu
02/01/2009	23	3.48	3.54	0.062	1.8	0.094	3.73	-17.6	-15.1	Fri
03/01/2009	23	3.42	3.47	0.046	1.3	0.106	3.68	-14.4	-16.6	Sat
07/02/2009	23	3.41	3.39	-0.020	-0.6	0.106	3.61	-16.1	-16.6	Sat
15/02/2009	23	3.40	3.41	0.009	0.3	0.143	3.70	-13.2	-11.6	Sun

15 kpl. yllä esitetyn kuvan korkeimpia kulutustunteja. Sarakkeessa real on toteutunut kulutus, sarakeessa predict ennusteen antama kulutus ja sarakeessa pred+2\*dev ennuste johon on lisätty hajonta kaksinkertaisena. Alle tämän jääneet ennusteen arvot on merkitty punaisella. T\_short on edellisen 8 ja T\_long edellisen 40 tunnin lämpötilakeskiarvo.



## Liite 11: Kaukolämpöasiakkaan kulutushuiput 2008-2009

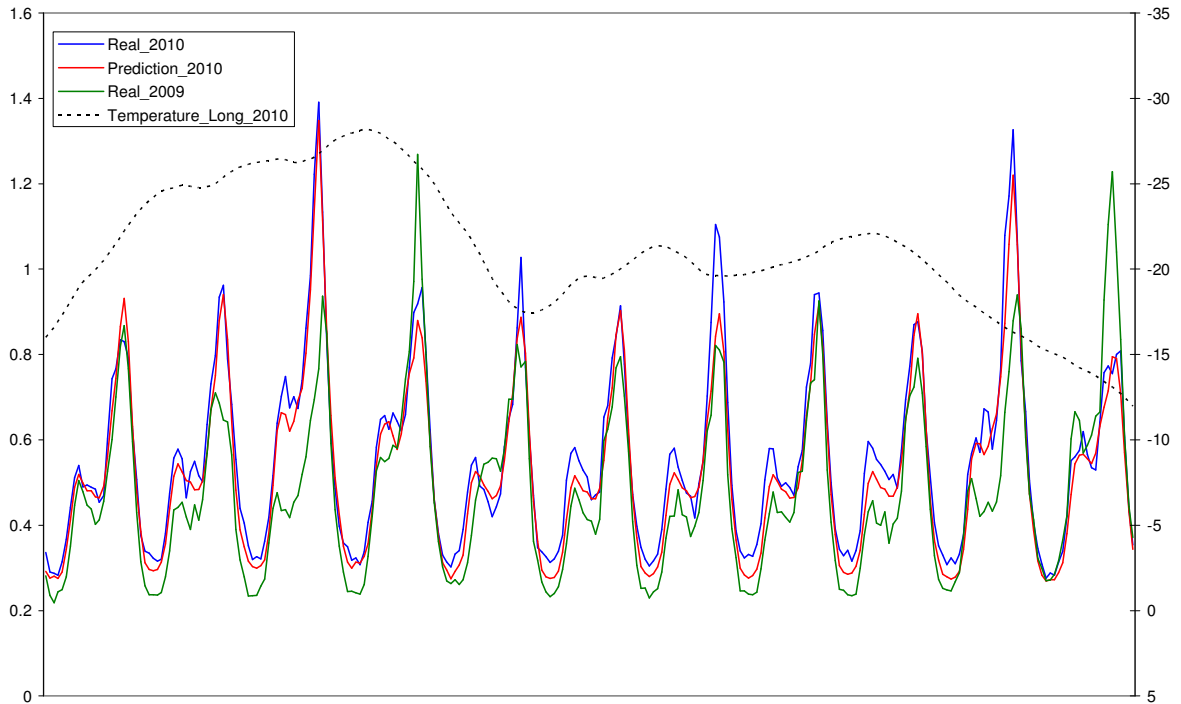


Keskimääräisen kaukolämpöasiakkaan tuntikulutukset järjestettynä alkuperäisen mittauksen suuruuden mukaan sinisellä sekä lialla ennusteen antamat arvot samoille tunneille. Käytössä oli perusjako kuuteen päiväryhmään.

#Date	h	real	predict	dif	dif %	stDev	pred+2*dev	T_short	T_long	Weekday
01/07/2008	0	1.38	0.22	-1.162	-84.1	0.132	0.48	16.9	18.4	Tue
20/12/2008	18	1.31	1.09	-0.225	-17.1	0.114	1.32	0.2	0.9	Sat
03/01/2009	18	1.31	1.21	-0.092	-7.1	0.114	1.44	-14.4	-16.6	Sat
07/02/2009	18	1.27	1.17	-0.100	-7.8	0.114	1.40	-16.1	-16.6	Sat
21/02/2009	19	1.26	0.97	-0.293	-23.2	0.098	1.17	-7.1	-7.0	Sat
24/01/2009	18	1.25	1.20	-0.056	-4.5	0.114	1.42	-9.9	-8.3	Sat
27/12/2008	18	1.24	1.13	-0.106	-8.5	0.114	1.36	-0.3	0.4	Sat
14/02/2009	18	1.24	1.12	-0.113	-9.2	0.114	1.35	-11.6	-10.9	Sat
31/01/2009	18	1.21	1.17	-0.039	-3.2	0.114	1.40	-17.9	-17.0	Sat
27/12/2008	19	1.21	0.99	-0.223	-18.5	0.098	1.19	-0.3	0.4	Sat
21/02/2009	18	1.19	1.09	-0.099	-8.4	0.114	1.31	-7.1	-7.0	Sat
17/01/2009	18	1.18	1.20	0.024	2.0	0.114	1.43	-15.2	-17.5	Sat
20/12/2008	17	1.14	1.02	-0.126	-11.0	0.091	1.20	0.2	0.9	Sat
13/12/2008	17	1.13	1.02	-0.117	-10.4	0.091	1.20	-0.3	-1.2	Sat
13/12/2008	18	1.13	1.08	-0.050	-4.4	0.114	1.31	-0.3	-1.2	Sat

15 kpl. yllä esitetyn kuvan korkeimpia kulutustunteja. Sarakkeessa real on toteutunut kulutus, sarakkeessa predict ennusteen antama kulutus ja sarakkeessa pred+2\*dev ennuste johon on lisätty hajonta kaksinkertaisena. Alle tämän jääneet ennusteen arvot on merkitty punaisella. T\_short on edellisen 8 ja T\_long edellisen 40 tunnin lämpötilakeskiarvo. 1.7.2008 klo. 0 on mitä luultavimmin mittausvirhe.

## Liite 12: Kaukolämpöasiakas 21-31.1.2010 kun jako kahteen lämpötila-alueeseen

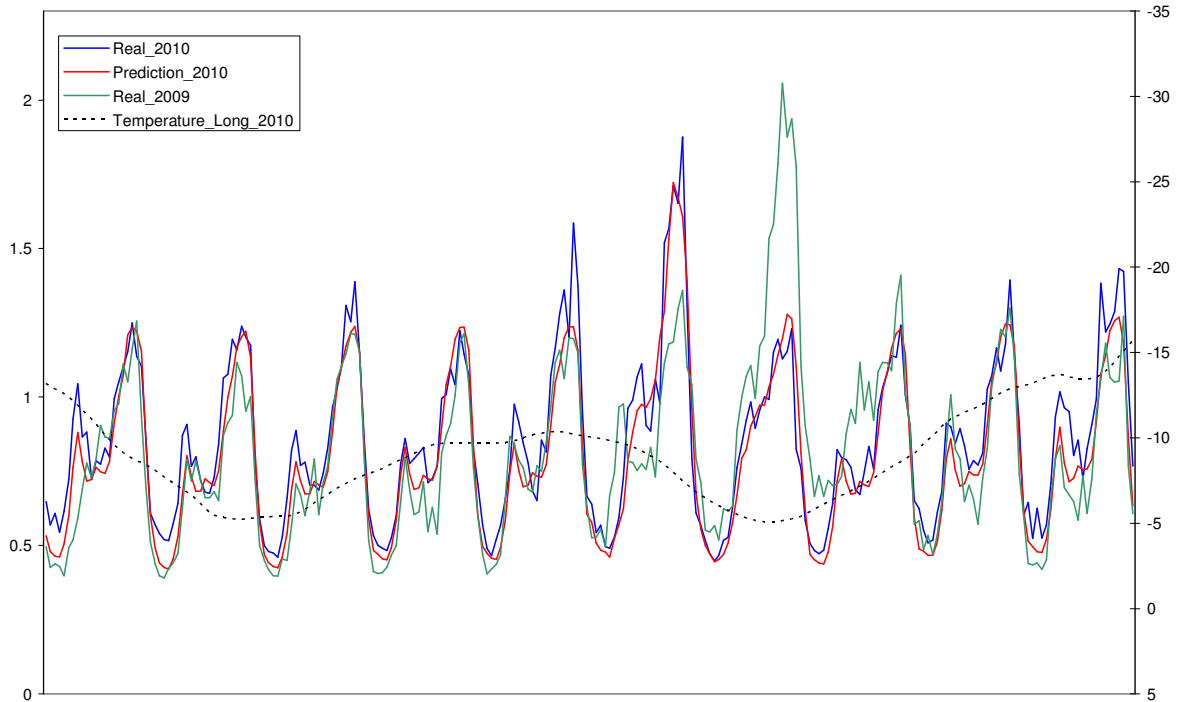


*Vihreällä vuoden 2009 mitattu kulutus, punaisella 2008–2009 datan avulla tehty ennuste kun käytetty 01/2010 lämpötiladataa ja sinisellä vuoden 2010 mitattu kulutus.*

#Date	h	real	predict	dif	dif %	stDev	pred+2*dev	T_short	T_long	Weekday
09/01/2010	18	1.46	1.30	-0.162	-11.1	0.114	1.53	-21.6	-23.9	Sat
02/01/2010	18	1.43	1.32	-0.118	-8.2	0.114	1.54	-19.5	-17.6	Sat
23/01/2010	18	1.39	1.35	-0.042	-3.0	0.114	1.58	-27.4	-26.7	Sat
30/01/2010	18	1.33	1.22	-0.106	-8.0	0.114	1.45	-15.2	-16.3	Sat
09/01/2010	19	1.30	1.12	-0.180	-13.8	0.100	1.32	-21.3	-23.7	Sat
23/01/2010	17	1.22	1.14	-0.082	-6.7	0.093	1.33	-27.2	-26.6	Sat
20/01/2010	19	1.21	0.91	-0.299	-24.8	0.080	1.07	-16.0	-14.4	Wed
02/01/2010	17	1.20	1.14	-0.056	-4.7	0.093	1.33	-19.1	-17.3	Sat
09/01/2010	17	1.19	1.13	-0.052	-4.4	0.093	1.32	-21.8	-24.1	Sat
30/01/2010	17	1.17	1.06	-0.113	-9.7	0.093	1.24	-15.2	-16.4	Sat
05/01/2010	19	1.16	0.92	-0.241	-20.8	0.080	1.08	-18.5	-19.1	Tue
23/01/2010	19	1.13	1.12	-0.019	-1.7	0.100	1.32	-27.5	-26.9	Sat
27/01/2010	18	1.10	0.84	-0.261	-23.7	0.067	0.98	-19.0	-19.6	Wed
02/01/2010	19	1.08	1.09	0.007	0.7	0.100	1.29	-19.9	-17.9	Sat
30/01/2010	16	1.08	0.87	-0.206	-19.1	0.058	0.99	-15.2	-16.6	Sat

*01/2010 keskimääräisen kaukolämpöasiakkaan huippukulutustunnit. Punaisella merkityissä ennusteen 95 % luottamustaso jäi alle 2010 mitatun kulutuksen.*

## Liite 13: Öljylämmittäjä 11-20.1.2010 kun jako kahteen lämpötila-alueeseen

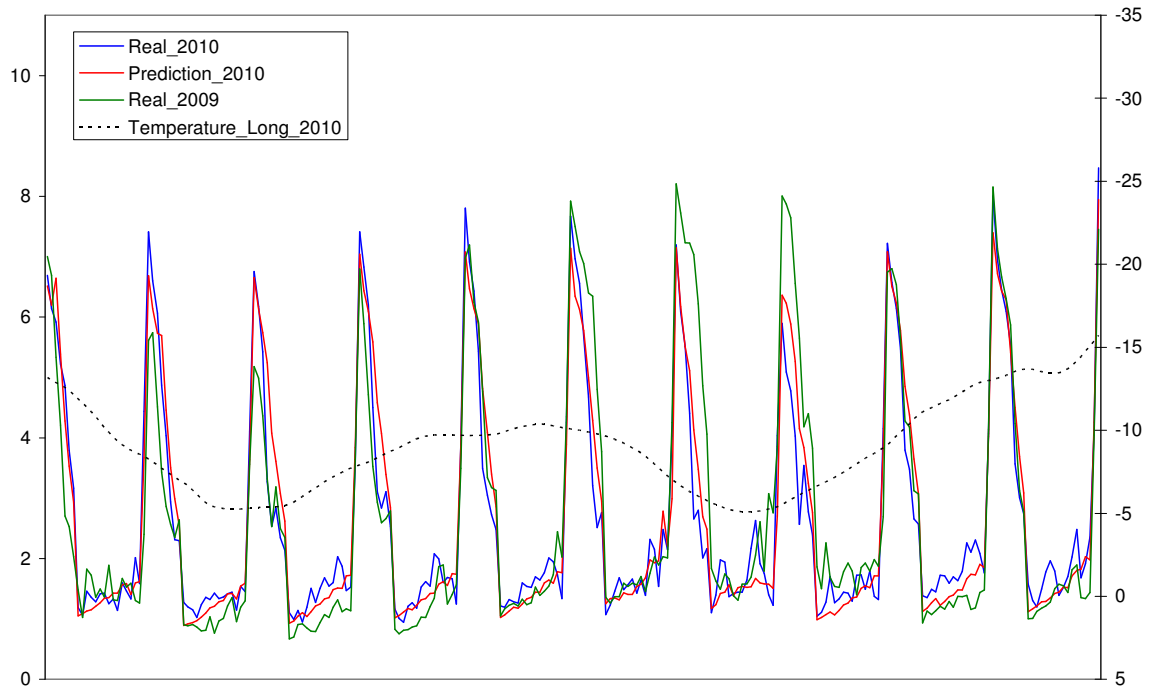


*Vihreällä vuoden 2009 mitattu kulutus, punaisella 2008–2009 datan avulla tehty ennuste kun käytetty 01/2010 lämpötiladataa ja sinisellä vuoden 2010 mitattu kulutus.*

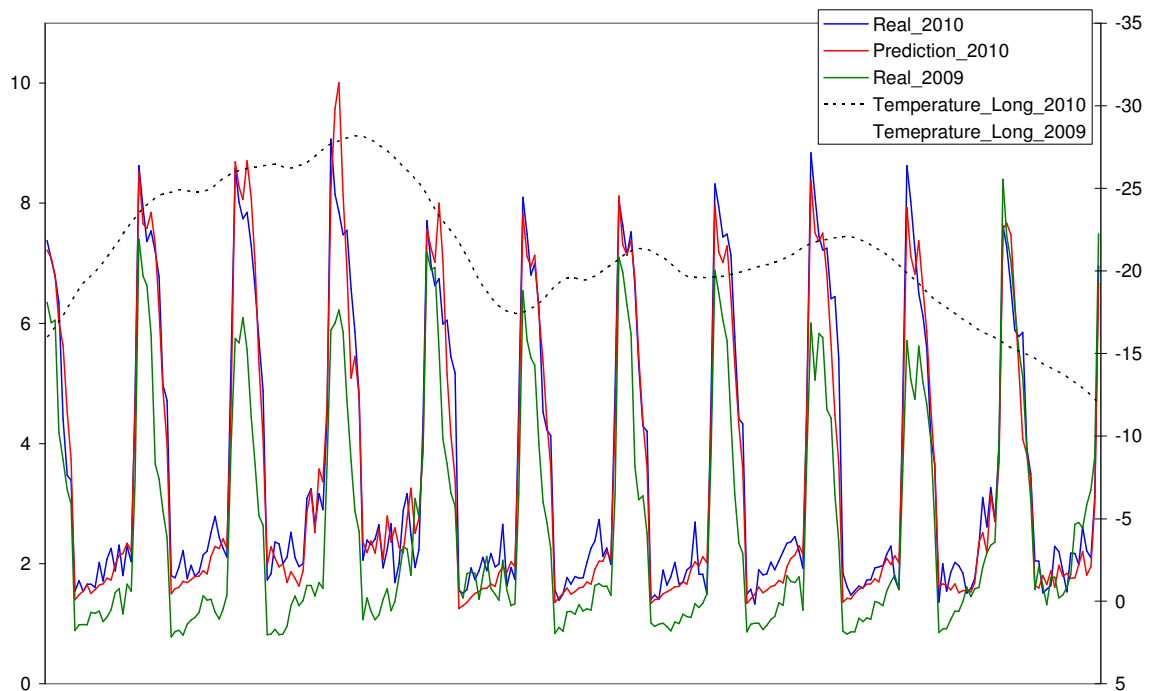
#Date	h	real	predict	dif	dif %	stDev	pred+2*dev	T_short	T_long	Weekday
23/01/2010	18	2.26	2.14	-0.129	-5.7	0.172	2.48	-27.4	-26.7	Sat
23/01/2010	19	2.10	1.84	-0.264	-12.6	0.183	2.20	-27.5	-26.9	Sat
30/01/2010	18	2.10	1.88	-0.220	-10.5	0.172	2.22	-15.2	-16.3	Sat
09/01/2010	18	2.10	2.04	-0.056	-2.7	0.172	2.38	-21.6	-23.9	Sat
23/01/2010	20	2.05	1.77	-0.277	-13.5	0.159	2.09	-27.6	-27.2	Sat
30/01/2010	19	1.92	1.72	-0.198	-10.3	0.183	2.08	-15.3	-16.2	Sat
02/01/2010	18	1.90	2.07	0.179	9.4	0.172	2.42	-19.5	-17.6	Sat
16/01/2010	20	1.88	1.61	-0.268	-14.3	0.159	1.93	-5.8	-7.5	Sat
05/01/2010	19	1.83	1.33	-0.502	-27.4	0.134	1.60	-18.5	-19.1	Tue
02/01/2010	19	1.82	1.83	0.007	0.4	0.183	2.20	-19.9	-17.9	Sat
09/01/2010	20	1.82	1.77	-0.053	-2.9	0.159	2.09	-21.0	-23.4	Sat
05/01/2010	18	1.80	1.31	-0.490	-27.2	0.120	1.55	-18.7	-19.2	Tue
09/01/2010	19	1.80	1.80	0.006	0.3	0.183	2.17	-21.3	-23.7	Sat
23/01/2010	17	1.77	1.86	0.087	4.9	0.138	2.13	-27.2	-26.6	Sat
02/01/2010	17	1.73	1.74	0.008	0.5	0.138	2.02	-19.1	-17.3	Sat

*01/2010 keskimääräisen öljylämmittäjän huippukulutustunnit. Punaisella merkityissä ennusteen 95 % luottamustaso jäi alle 2010 mitatun kulutuksen.*

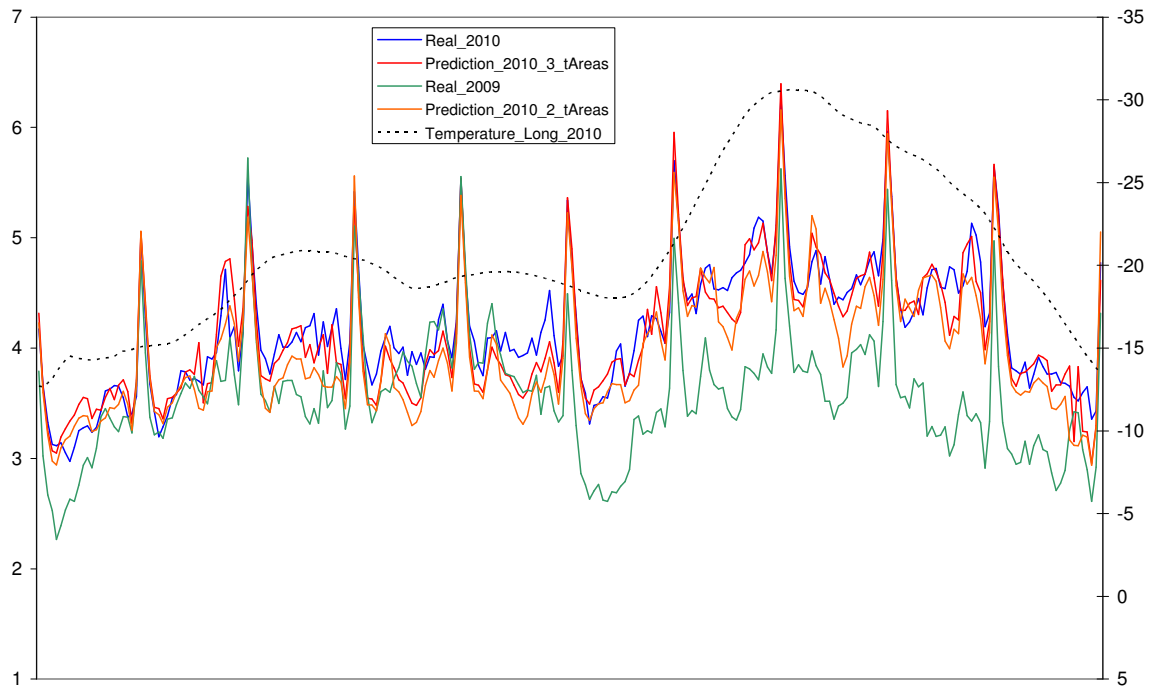
## Liite 14: Varaava sähkölämmittäjä 11-31.1.2010 kun jako kolmeen lämpötila-alueeseen



*Ensimmäisessä kuvassa jakso 11-20.1.2010 ja toisessa 21-31.1.2010.*



## Liite 15: Ilmalämpöpumpua käyttävä suora sähkölämmittäjä 1-20.1.2010



*Yllä jakso 1-10. ja alla 11-20.1.2010. Punaisella 3 ja oranssilla 2 lämpötila-alueen ennuste.*

